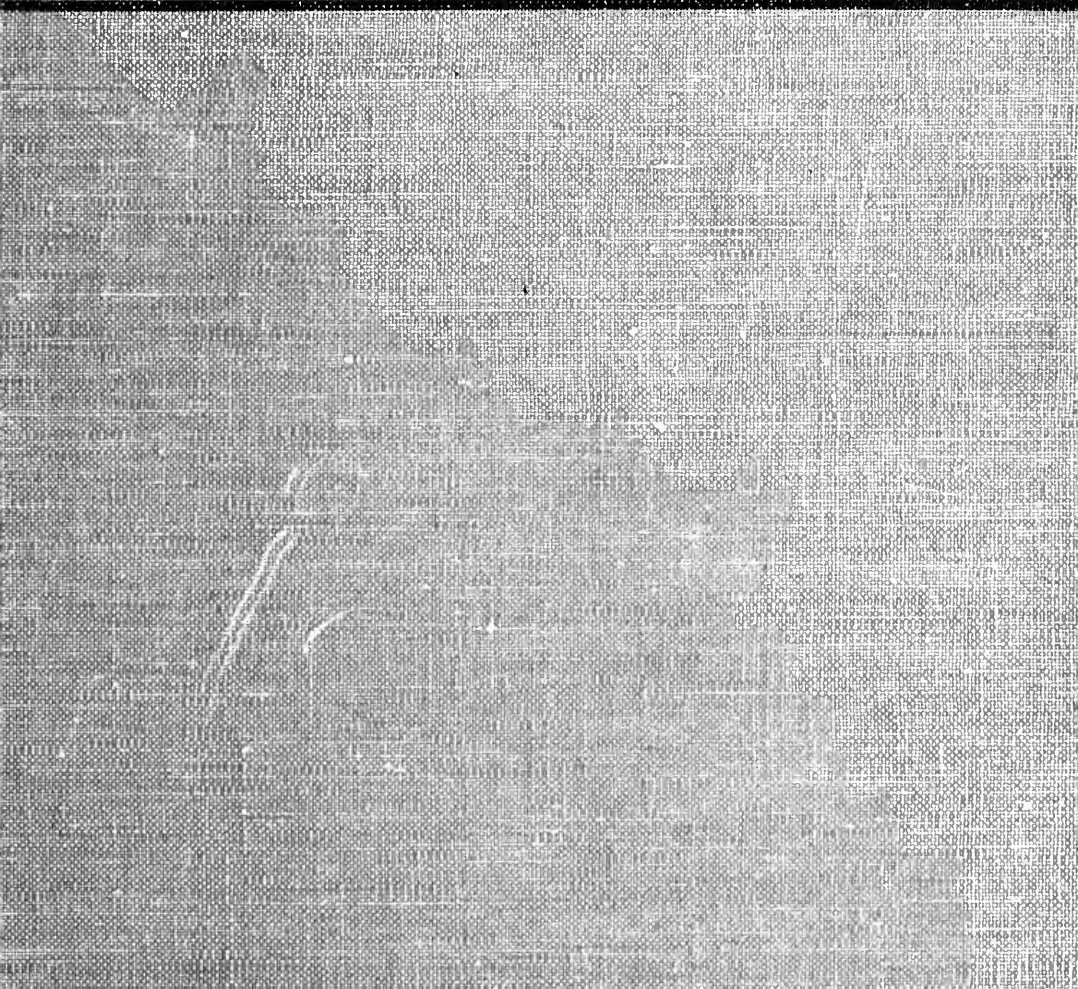# BIOPHYSICAL
RESEARCH METHODS

SHELF LIST

0 0301 0012002 8

BIOPHYSICAL RESEARCH METHODS

BIOPHYSICAL
RESEARCH METHODS

Prepared by a Group of Specialists under the Editorship of
FRED M. UBER
Navy Electronics Laboratory, San Diego, California

PREFACE

This collective volume is intended to serve as a stimulating but critical guide to that rapidly growing group of scientists who must resort to physical methods of research for the solution of biological, medical, and agricultural problems.

Each chapter provides an authoritative orientation with respect to one general research method. To achieve this end, each chapter covers the following points: (1) the fundamental principle of one method, its underlying assumptions, and perhaps a simple mathematical outline of the theory involved; (2) the types of problems for which the method offers promise of a unique solution or a helpful approach; (3) the demands made on apparatus, materials, and technical skill; (4) the limitations of the method in some detail; and (5), briefly, its outstanding accomplishments to date. After reading any chapter, a scientist should be able not only to appraise the potentialities of its method with reference to his own biological problem, but also to understand the limiting factors that must be given recognition in the proper design of critical experiments.

The book should prove useful in a wide variety of advanced courses, particularly for collateral reading assignments aimed at the mature student. Each chapter can be read as an independent unit without regard to its position in the volume.

References to commercial sources of equipment and materials have been included by the several authors as a convenience to the reader. Listing of a product and its source does not necessarily constitute an endorsement nor does failure to list a product indicate its inferiority in any way.

My thanks are willingly expressed to the several authors for their kind cooperation at all stages of the work. Others who offered helpful suggestions during the planning stage include: Dr. L. R. Blinks, the late Dr. S. C. Brooks, Dr. C. S. French, Dr. D. R. Goddard, Dr. W. M. Stanley, Dr. Otto Stuhlman, and Dr. Maurice Visscher. Most of the editorial work was accomplished while I was Professor of Physics at Iowa State College. Consequently, I feel most indebted to the Department of Physics at Iowa State College

for facilities and secretarial assistance; and particularly to Professors Jay W. Woodrow, Gerald W. Fox and Percy H. Carr for their encouragement. At the Navy Electronics Laboratory, the cooperation of Dr. R. J. Christensen has been very helpful.

CONTRIBUTORS

HAROLD F. BLUM, *National Cancer Institute and Department of Biology, Princeton University, Princeton, New Jersey*

DAVID R. BRIGGS, *Division of Agricultural Biochemistry, University of Minnesota, St. Paul, Minnesota*

HOWARD J. CURTIS, *School of Medicine, Vanderbilt University, Nashville, Tennessee*

EARL W. FLOSDORF, *F. J. Stokes Machine Co., Philadelphia 20, Pennsylvania*

JOHN W. GOWEN, *Department of Genetics, Iowa State College, Ames, Iowa*

L. H. GRAY, *Radiotherapeutic Research Unit of the Medical Research Council, Hammersmith Hospital, London W. 12*

EARLE C. GREGG, JR., *Department of Physics, Case Institute of Technology, Cleveland 6, Ohio*

JAMES D. HARDY, *Department of Physiology, Cornell University Medical College, New York 21, N. Y.*

L. V. HEILBRUNN, *Department of Zoology, University of Pennsylvania, Philadelphia 4, Pennsylvania*

JAMES HILLIER, *RCA Laboratories Division, Radio Corporation of America, Princeton, New Jersey*

MAX KLEIBER, *College of Agriculture, University of California, Davis, California*

E. G. PICKELS, *Specialized Instruments Corporation, Belmont, California*

LAWRENCE R. PROUTY, *Department of Physiology, Cornell University Medical College, New York 21, N. Y.*

OSCAR W. RICHARDS, *Scientific Instrument Division, American Optical Co., Buffalo 15, New York*

FRED M. UBER, *U. S. Navy Electronics Laboratory, San Diego 52, California*

ADOLF F. VOIGT, *Institute for Atomic Research and Department of Chemistry, Iowa State College, Ames, Iowa*

CONTENTS

ix

BIOPHYSICAL RESEARCH METHODS

AVOID FRUITLESS EXPERIMENTS

Fred M. Uber, *U. S. Navy Electronics Laboratory*

Few scientists, even on a holiday, would undertake deliberately to conduct a meaningless experiment.   But most critical scientists will

subscribe, I believe, to the assertion that numerous research articles are either meaningless or destined to be fruitless. Some published evidence exists to support this statement (see Sect. H) and if scientists were less courteous to their fellow workers, there probably would be much more.

The freedom to exercise scientific curiosity experimentally in a laboratory is a priceless heritage of modern science. This freedom, unexploited in ancient times and even denied during the middle ages, should be guarded zealously; it should not be jeopardized by careless indulgence and irresponsibility on the part of some investigators. To be valuable, contemporary research must place increasingly greater stress on the careful analysis, design, and often on the co-operative execution of experimental programs. To point a finger at some of the pitfalls which needlessly embarrass too many naive experimentalists is the aim of this introductory chapter. The writer is conscious of some of its limitations, but would welcome constructive suggestions from any source for its improvement. Specific examples and illustrations particularly are solicited. It is a pleasure to acknowledge my indebtedness to several of the sources mentioned in the bibliography and especially to an inspiring lecture by T. S. Hamilton (22).

A. APPROACHES TO IMPORTANT DISCOVERIES

To prescribe what must be done to avoid meaningless or fruitless experiments is relatively easy compared to the task of outlining a procedure for selecting the most meaningful and important problems. Were this not true, science would be much further advanced. No attempt will be made in this chapter to outline any such procedure, for reasons which probably are or will become obvious to the reader very soon. It is felt, however, that the time which could be saved by avoiding fruitless experiments might result in appreciably increasing the number of important discoveries. Some widely held, but conflicting, viewpoints on how to make fundamental discoveries will be presented now for the purpose of background to the fundamental research situation.

1. Accidental Approach

Not infrequently in the past, important scientific discoveries have been made by workers who were not trained to look for them. The

idea is rather widespread that great discoveries are largely accidental in nature and that they may happen to almost anyone. The accidental theory of success in research has been strikingly epitomized in a recent remark by an American beer baron. With business booming, the baron had permitted himself the luxury of subsidizing a capable microbiologist, but subject to an annual appropriation. At a year-end conference on the budget, to which the laboratory superintendent had invited the microbiologist, the question of a renewed grant was up for consideration. But the baron quickly settled the matter. "Keep him on," he said with a flourish, "you never know when he might stumble onto something."

To ascribe the tremendous achievements of modern science to an unending series of accidental discoveries not only is placing "Lady Luck" on a very high pedestal but also is unfair to professional scientists. However, whereas an occasional important discovery *may* be accidental, most meaningless experiments are the result of inadequate planning and/or careless execution. While it may be professionally embarrassing for some to admit that an epochal discovery has been made in a backwoods attic laboratory by a novice with crude equipment and a high school education, it would seem downright disgraceful to be forced to acknowledge that many research experiments are not adequately designed by supposedly well trained investigators.

Another type of accidental approach is involved in the contention that experiments that are wrong ultimately lead to the great advances in science. The thought here is that the large amount of careful work required to prove an experiment incorrect results itself in fundamental contributions. This becomes secondarily a hybrid situation which includes incidental elements of the type discussed in the next section. However, the provocative incorrect experiment must be considered accidental—unless eventually efficiency experts deliberately publish such articles in order to stimulate research.

2. Incidental Approach

Important discoveries in science in recent years usually have been made by qualified investigators deliberately engaged in experimental research. Oftentimes it has happened that the important advance was merely a by-product, however, of the original problem under investigation. To this extent, then, the discovery is incidental. If the reader prefers to call it accidental, I believe he will concede that

it is a type of accident that does not happen to just anyone. It should
occur most frequently to investigators who are alert to the unusual
and who are able to perceive quickly the possibilities in the unexpected
observation (*cf.* Fig. 1). The fact of X rays, for example, represents a
phenomenon that might have been discovered by anyone of perhaps a
few hundred experimental scientists at the time, but almost incon-

"*Now, I sort of lose the gist of it here.*"

Fig. 1.   An unexpected observation often leads to significant advances. (Cour-
tesy *The New Yorker.*)

ceivably by one of the billion contemporary laymen. According to
this approach, great discoveries are favored by the hard work of
numerous careful observers who are also resourceful.

3. Deliberate or Direct Approach

A considerable part of the motivation of research workers is
doubtless the hope of making a really important discovery. In some
scientists this purpose expresses itself as a deliberate, direct attempt.
An investigator in this category analyzes every proposed problem
primarily on the basis of its possible importance. Much can be said
in favor of such a critical attitude. At the least, it prevents a worker
from embarking on programs in which he has too little faith, the type

undertaken merely with the hope that something worth while will result. The observer who is *looking for* something specific seems to have a greater chance of seeing it than an observer who is *just looking*. An investigator with a strong conviction that the natural world operates on simple laws has a stronger incentive to work doggedly to discover them than an individual who lacks faith in their existence. Absolute devotion to a particular problem, however, must surely end in failure in a great many cases; the number of relatively important discoveries are but a fraction of one per cent of the number of experimental attempts.

The striving for important discoveries, for the "order-of-magnitude" advances, therefore carries a high degree of risk. Some scientists are much better situated than others with respect to assuming this extra hazard. Apart from the mental adjustment and the intestinal fortitude demanded, there is usually an economic or timing factor. This is clearly true with most problems for graduate students. Beginners generally feel the need for early successes, but it can hardly be said that they gamble less than their older colleagues. The student's gamble, however, is often on his professor's judgment rather than deliberately on the problem.

This type of approach doesn't seem well suited to workers who need the frequent stimulation which results from minor successes. The great-gamble type of experiment is not necessarily devoid of by-products and incidental data which are of themselves valuable, provided the time and effort are expended to make them so. But the latter procedure detracts at once from the effort that can be put on the main purpose of the problem or on the next logical attempt at an important discovery. Perhaps few scientists are *entirely* free to pursue their own desires in the matter.

4. Organized or Controlled Approach

In commenting on the influence of Francis Bacon on the scientific revolution, Mees (4, p. 81) has stated that "Bacon over-estimated the ease with which scientific knowledge can be obtained, and he fell into an error in which he is followed by many today—the error of believing that scientific research can be organized like an engineering project and that the way to make scientific discoveries is to plan to make them." A great many scientific contemporaries share this view with Mees and are distrustful of too much "direction" of scientific research programs. President Conant of Harvard University has been quoted

as saying, "There is only one proved method of assisting the advancement of pure science—that of picking men of genius, backing them heavily, and leaving them to direct themselves."

It should be emphasized that we are discussing approaches to *important* discoveries—without knowing in what direction they lie. The danger inherent in an organized or controlled approach is simply that of concentrating too much effort blindly in too few directions, thereby possibly missing altogether a really fundamental discovery. A strategic break-through on the research frontier would seem to become increasingly probable when many independent investigators are engaged in the search, each on his own initiative. Furthermore, the morale and efficiency of research talent might be lessened tragically by any large scale attempt at organized planning. Now that research scientists have achieved a very considerable measure of success, it is to be expected that efficiency experts will hover about their laboratories to tell them how to plan their experiments—if not, indeed, to control their investigations completely.

B. ANALYZE THE PROBLEM

Scientists resort to experiment when questions or problems arise which cannot be disposed of satisfactorily in any other way. Problems may be relatively trivial so that their experimental solution can be found in a matter of hours and at small cost in terms of time and money; or they may be exceedingly complex so that only a very extensive investigation requiring years of effort can conceivably result in a satisfactory elucidation of the difficulty. A current example of the latter type is the cancer problem. A broad problem such as this can be approached from many angles. There may even be more than one satisfactory solution. I believe that most experimental scientists expect without doubt that a solution will be found eventually. In the meantime numerous aspects of the problem present themselves and it is clear than an unlimited number of experiments could be performed to collect factual data. When ultimately an acceptable scientific answer is found, much of this experimentation will be regarded in all probability as having been quite useless. The bulk of the attempts will not have served any critical function. But how can such meaningless investigations be avoided? I believe that a careful analysis of the problem would do much to eliminate a large percentage of these fruitless experiments.

What constitutes a *careful analysis* of the problem? So many

factors are involved that there is no single well marked groove or alley that can be followed blindly. A complete solution of a broad problem may require step-by-step progress through the various stages of what is often called the "scientific method." These stages have been enumerated recently by Northrop (*3*, p. 28) as follows:

(*a*) Discover the basic theoretical root of the problem.

(*b*) Select the simplest phenomenon exhibiting the factors involved in the difficulty.

(*c*) Observe inductively these relevant factors, either by the method of observation, the method of description, or the method of classification.

(*d*) Project the relevant hypotheses suggested by these relevant facts.

(*e*) Deduce logical consequences from each hypothesis, thereby permitting it to be put to an experimental test.

(*f*) Clarify initial problem in the light of verified hypotheses.

(*g*) Generalize solution to the problem by means of a pursuit of the logical implications of the new concepts and theory with respect to other subject matter and applications.

Individual investigators often pursue a very limited phase or aspect of a broad problem; for example, a number of scientists have devoted a lifetime of research to the purely descriptive phase of the cancer problem. This is most essential, but it constitutes only one part of the whole. In any event, an investigator should realize clearly how his individual research may contribute toward a general solution of *the* problem. The failure to view one's experimental work in the light of the broad over-all situation may result in the prosecution of numerous fruitless experiments.

1. Look for Basic Difficulty

You may wish to refer to the theoretical basis of a problem as its heart, its core, or perhaps its quintessence. That a problem exists at all would seem to indicate that a fundamental difficulty of some kind is present. The object is to uncover it, to understand its nature, in order that a solution may be obtained. No attempt will be made to set forth rules as guideposts to the heart of a research problem. Even hints as to how to grope for it cannot be given, but prolonged and serious reflective thought about the nature of the difficulty is highly recommended. A strictly armchair approach probably will not be sufficient. Many problems cannot be analyzed adequately until after extensive preliminary observations of a descriptive type are available. Not infrequently, exploratory investigations of the experimental

variety must be conducted before it is profitable to try to lay bare the
heart of the difficulty.   This information may, of course, have been
published already by others and thus be accessible.

In the process of arriving at the theoretical basis of a problem,
there may be an intricate interplay between fact and fancy, between
ideas and experimental data, between hunches and lucky observations,
and between intuition and blunders. Most scientists do not record
the tortuous paths traversed in the pursuit of their discoveries. The
fact that some famous scientists have been responsible for several
basic discoveries discounts the cynical view that scientific progress is
purely an accidental process.

External stimulation of the thought processes should not be
neglected in an effort to achieve an insight into problematical situa-
tions. Apart from the stimulus that derives from germane conversa-
tion and from browsing in likely fields of literature, a conscious effort
should be made to broaden one's contacts with other, and perhaps
remote, domains of scientific endeavor.

2. Focus Attention on Relevant Facts

"It is to be noted that it is the analysis of the problem which
provides the criterion for selecting out of the infinite number of facts
in the world the few that are relevant," to quote Northrop (*3*, p. 34).
To proceed willy-nilly to collect irrelevant facts is not regarded as
good scientific procedure, although it may occasionally result in a
solution.   There is also a possibility that the necessary and relevant
facts already exist in the literature.   Even after an armchair analysis
has led one to the basic root of the problem, there may be several
other factors that will have an important bearing on whether a given
individual should undertake an experimental investigation. Some of
these will be discussed in later sections.

To be most fruitful, experiments should be undertaken only to
solve *bona fide* problems.   Unless based on a genuine problem, an
investigation lacks purpose and directive force.   Only chaos can
result from an attempt to record all possible facts.   Even Charles
Darwin, whose chief service to science was the production of a mass of
descriptive evidence that evolution has occurred, has said, "How odd
it is that anyone should not see that all observations must be for or
against some view, if it is to be of any service."   It is the problem that
determines, for example, with what precision measurements are to be
made and how extensive an experiment is to be.   The discontinuous

nature of experimental data demands that some appropriate interval
of time be chosen as a guide in making observations. The value of
this interval must be determined by the nature of the problem itself.

In making observations, the scientist selects some of the facts for
attention; he does not attempt to record all of them. Scientific facts
represent relatively only a very small number of all the facts that
could be observed. Although scientists deliberately choose certain
facts for observation, the selection must be made on the basis of proper
criteria and in an impartial manner according to the unwritten code
called *scientific honesty*. The choice is not one of mere caprice nor can
it be purely arbitrary in order that the facts may fit into some pre-
conceived scheme. To be accepted as fact, experiments must be re-
producible by others.

The ability to distinguish readily the critical factual items in an
experimental situation is the mark of a capable scientist. Although
good scientists must be opportunists in the sense of making the most
of unexpected observations, they must also simulate bloodhounds and
not be distracted constantly by irrelevant facts of an inconsequential
nature.

3. Discard Meaningless Questions

Scientific experiments are undertaken in order to answer questions
concerning nature. To state that an experiment is meaningless
either implies failure in securing an unequivocal or satisfying answer
to the question, or it suggests that the question itself is without
meaning. In a philosophical sense, one can defend the thesis that
seriously propounded questions concerning nature always possess
meaning. Only when a specific point of view is adopted does it become
possible to define questions as meaningless or to assign a scale of
values to them. One such point of view has been expounded by
Bridgman (*1*, p. 28), who classifies questions as meaningless unless
they can be answered by means of "operations." For instance, it
means nothing to ask whether a star is at rest or not. To laboratory
workers, "operations" connote experimental manipulation and ob-
servation. From this standpoint, for example, the question of whether
there was once a time when matter did not exist possesses no meaning.
Presumably in the same category is the question whether a rabbit and
a mouse experience identical sensations to the color red. But all
questions that could be answered in terms of operations would be
meaningful.

In a narrower sense, many potentially meaningful questions may be meaningless at a given time or to a particular worker because methods or equipment might not exist with which he could conceivably perform the necessary "operations." An example: Are there mountains on the other side of the moon?

A recent discussion of what constitutes meaningful questions with respect to experimental science has been given by Churchman (*5*), who elaborates and extends the concept outlined by Bridgman.

4. Appraise Relative Importance of Problems

Individuals are constantly appraising the relative values of the several experimental sciences and of the more limited fields of investigation within specific branches of research activity. It is only natural that serious consideration be given to the possible significance of a proposed research program and its importance as related to the broader aspects of science and other human endeavor. The task is not an easy one. If history has taught any lesson clearly concerning the ultimate value of research to society, it is the fact that such values are unpredictable in advance. This results in part from the unexpected by-products of experimental studies but in large measure from the autocatalytic nature of cumulative scientific knowledge. Hence, a conscious steering of basic research by society would seem to constitute an unwise procedure, if not an unconscious boomerang. Even scientists, far from claiming omniscience either as individuals or as a group, are not qualified to render *a priori* decisions regarding the eventual value or meaning to society of scientific questions.

Even though society cannot render a verdict in advance as to the ultimate significance of scientific questions, it can often appraise the experimental results. For example, an experiment can be relatively fruitless to society if it constitutes merely a repetition of an earlier published finding or if it is not made available to the public at all, as is true of much commercial research, thus leading to further duplication of effort. Where results are not readily comparable to related investigations and where experiments are not carried to satisfactory and unambiguous conclusions, they may likewise possess little meaning to others. The basic reason for the abundance of meaningless experiments is simply the failure of numerous experiments to supply an unequivocal answer to the question. Results that are presumably satisfying to the investigator himself, at least in some degree, are often unconvincing to scientists generally.

5. Choose Initial Problem Wisely

For the individual contemplating a scientific research career, the choice of a first problem is a matter of extraordinary importance. This first problem often determines the course of a life's work. If it is an unfortunate choice, it may even result in the individual's turning to other fields of endeavor, perhaps forsaking research entirely.

All too often the selection of an initial problem represents the culmination of a haphazard procedure. Not infrequently, the problem derives as a consequence from the choosing of a research professor. Although this method has often worked out to the benefit of the beginner, still it should not be accepted as a routine matter—the percentage of failures is too great. Instead, a very real effort should be made to appraise the relative merits of the various problems under consideration on the basis of specific criteria. The final decision should be deliberate on the part of the individual; it should not be made for him by default or otherwise. In addition to the thoughts expressed under the various subdivisions of this chapter, many of which have a bearing on the selection of a problem, a few comments especially applicable to the beginner may be helpful. Apart from his possible immaturity and lack of experience, a beginner usually suffers also from time limitations. Therefore, a problem should be attractive not only for its own sake but also for its prospect of completion within a reasonable and rather definite length of time.

The following excerpts from an article by Livingston (23) are very germane and cover some suggested criteria for judging a proposed research problem for the beginner:

"There are still those who maintain that any piece of scientific investigation carefully done and published must of necessity bear great fruit in future years, but such views are usually met with in those who do not seriously attempt to keep up with the progress of the current literature of their science.

"The problem chosen should be circumscribed, definite and specific. At the same time, it must be appreciated how this particular problem is related to other similar specific questions, the whole series covering some broad and general field. It frequently happens that a problem which attracts and fascinates a graduate student is far too broad to be rationally attacked, sometimes the mere breadth constitutes an attractive feature and throws a false glamour over the entire proposition. Such a question should be separated into partial questions and these attacked singly. The attempting of too broad and, hence, too indefinite a problem in his earlier years of research has worked lasting injury upon many a man of science.

"The ideal problem for a beginner should be capable of statement, on *a priori* grounds, in the form of several alternatives; all the logically possible answers to the questions may be advantageously erected into hypotheses, and these may be tested in order. This sort of a problem conduces to logical thinking and must leave its sterling mark upon the mind in later years. Furthermore it is economical of energy and time, and the end of the chosen piece of work is more or less clearly in view at the beginning. To bring a problem into this condition requires, of course, a large amount of thought at the outset. . . .

"The satisfactory problem must, of course, be capable of experimental treatment with the knowledge and facilities which are available. The beginner should not be called upon to devote too much time and energy to the devising of methods and the obtaining of apparatus. If he be misled in this he almost surely becomes more interested in the methods than in the results obtained by their employment. This does not imply that the methods to be used should all be familiar to the worker at the start, only that they should be accessible in the literature, so that he need not actually devise them.

"Apparent importance to the science as a whole is a very important criterion in our series. For the best results in all ways, the selected question should be one that interests both the theoretical and the practical worker. . . . The question of the theoretical importance of a given problem is not so easily settled as is that of its practical weight; it requires something of a prophet to judge rightly in this regard. A good way to attack this question is to ask, will any chapter of the science (as it stands at present) be fundamentally altered by the proposed study? . . . A superficial study of a little-known relation is often as important in the development of a science as is a research upon the details of a better-known and already more thoroughly analyzed phase. Such superficial studies are the work of pioneers; they are adapted only to the exceptional beginner in research."

6. Pursue Type of Research That Comes Naturally

Within the framework of the general aim that dominates all scientific work, there may be various subsidiary purposes peculiar to individual types of investigation. Some specific examples are:

(*a*) To test the limits of application of a general theory.
(*b*) To explore a new field for its possibilities.
(*c*) To create an instrument of measurement.
(*d*) To develop or improve some experimental method.
(*e*) To determine constants with a high degree of precision.

A worker who excels in doing experiments of the highest accuracy might not be very successful in exploratory investigations which do not utilize this special talent fully. Scientists accustomed to making

all observations personally often feel uncomfortable in administrative research positions where such duties are delegated to assistants. It seems sensible to devote one's efforts in a manner that will prove most fruitful eventually, especially since there is great need for all types of scientific activity.

The beginner must face the question as to whether he should engage in scientific research at all. Natural talent in this direction should be the criterion. Some help in arriving at a decision may be obtained from aptitude tests. Such tests may give an indication not only of general research aptitude but also of specific types of investigation in which the individual is most likely to succeed. It seems increasingly clear that many types of research will be done in the future by co-operative teams or groups. If an individual is not inclined to be co-operative, he should select his research field carefully in order to avoid problems requiring this quality.

C. REFRAIN FROM UNDUE REPETITION OF WORK OF OTHERS

1. Keep Abreast of Current Developments

Scientific literature has increased in bulk to the point where it is becoming increasingly difficult to avoid entirely the unwitting repetition of experiments. At the same time it has become perhaps more important to devote scientific energies to the new and unexplored. It is always disturbing when exciting and seemingly original research ideas turn out upon further inquiry to have been studied by others years ago, yet the disappointment should be easier to bear if it occurs before rather than after the unknowing duplication of previous work.

Where research programs involve substantial amounts of time and money, a fairly extensive library search relative to past achievements along the same line should be considered as certainly justified if indeed not obligatory. Whether exhaustive searches into the literature are conducted depends both on the idiosyncrasies of the individual and on the ease with which such information can be obtained. Some investigators rarely enter a library, while others may spend too much time in reference work. Unless a genuine attempt is made to search the literature for past achievements, it is difficult to justify a research program. To repeat from ignorance experiments already satisfactorily performed and reported is not conducive to scientific progress.

The time is long past when one could hope to keep abreast of the

advances in even his own field of research by browsing occasionally in a library or by depending on the receipt of reprints from the workers themselves. Accordingly, one should plan to make systematic use of some or all of the following categories of assistance with reference to his special field of research (*19*, p. 74): (*a*) index journals; (*b*) abstracting journals; (*c*) annual reviews and yearbooks; (*d*) recent advances series on specific topics; and (*e*) review journals. For those without experience in library research, a list of "guides" to the literature in the various sciences is included in the bibliography (*13–18*). These contain helpful suggestions as well as source lists of specific reference materials for the individual sciences.

In connection with reading the older literature, a word of caution may be helpful, particularly with reference to numerical data. Over the years, a marked change has taken place in the value of the numerical physical constants that enter into various computations. Hence, recalculation is in order before making comparisons with many of the earlier results. Common usage for words and symbols also change. A good example to illustrate the point is the recently suggested change in nomenclature involving the use of L and D instead of l and d for labeling the amino acids and sugars (*25*). The opposite name from the one formerly employed results in some cases, to the confusion of the uninitiated. Nomenclature changes for the genus and species of microorganisms, especially bacteria, are not infrequent and are likewise confusing.

There is no unique way of keeping in satisfactory contact with research currently in progress, particularly where investigative activity in an experimental field is world-wide. With respect to investigations under way within one's own country, valuable information can be gleaned from attendance at scientific meetings and perhaps from correspondence, but, with the exception of very limited fields, the coverage will be inadequate at best. There has been an increasing tendency to organize scientific workers into small groups on the basis of their interest in specialized fields of research. Recent examples of this are such organizations as the Electron Microscope Society, the Society for X-ray and Electron Diffraction, and the Phycological Society. Informal symposia on specialized topics of active research interest are becoming increasingly frequent and should go far toward keeping scientific specialists up-to-date. The proceedings of such symposia are often published, but only after an appreciable lapse of time. There seems to be some danger that the great effort required to

keep informed may limit seriously the total research accomplishments.

2. Shun Negative Experiments

Most investigators at one time or another have repeated experiments previously reported by others, usually because it was felt that the conclusions of the original experiments were incorrect. Occasionally, however, the importance of a particular finding is such as to render its early confirmation highly advisable. Some investigators hesitate to attempt any repetition of earlier experiments as a matter of policy, feeling that a confirmation contributes nothing new and that a contradiction cannot be reported unless it is based on an exhaustive and extremely careful study. Since positive experimental results normally carry much more weight than negative, a greater than usual burden of proof is therefore placed upon the investigator who attempts to establish negative results in opposition to a previously reported positive finding. Nevertheless, it is very important to discredit erroneous findings at the earliest opportunity.

The repetition of an experiment for the express purpose of disproving previously published results is not advisable unless a worker is prepared to repeat the investigation in the painstaking and thorough manner necessary to convince the skeptical. Serious thought should be given in advance to the essential factors that constitute a satisfactory negative experiment. On what critical points must the eventual decision, regarding the facts to be accepted, be determined?

In the history of science there have been a number of instances in which established investigators in a field have failed to acknowledge their own published mistakes. This constitutes a fault that is very difficult to condone especially in a scientist. Since other workers are prone to amass more and more evidence confirming the original mistake, the meanness of one man can thus lead to a great number of otherwise needless and perhaps completely fruitless experiments.

D. RECOGNIZE EXPERIMENTAL LIMITATIONS

1. Theoretical Considerations

All research methods are subject to certain limitations and the realization of this fact can often result in a saving of experimental effort that can then be directed into more profitable channels. These

boundaries may be either of an experimental or of a theoretical nature.
Some are specific and apply to narrow regions or to particular tech-
niques; others are general and may encompass an entire field of
scientific investigation. An example of the latter is contained in the
broad generalizations of the first and second laws of thermodynamics.
The first law expresses the well known principle concerning the con-
servation of matter and energy. The second law is something of an
oddity in that it is a negative statement which tells what cannot be

**"He swore he had a pedigree a mile long, and I,
like a fool, believed him!"**

Fig. 2.   An experimental scientist should not depend on hearsay.   (Courtesy
Liberty Magazine.)

done and is therefore not subject to experimental confirmation. For
centuries, however, individuals have tried diligently but vainly to
disprove it by their ingenious attempts to attain perpetual motion, in
other words, a method of obtaining something for nothing. Such
people are now referred to as "perpetual motion cranks." They have
been ignored by orthodox physicists for a long time, but their recurrent
stories still make newspaper copy.

Of more restricted application but yet an excellent illustration of a
boundary imposed by physical theory is the theoretical resolving
power of the optical microscope. This somewhat arbitrary limit has
been confirmed repeatedly by experiment and tested so thoroughly

that it seems an inexcusable waste of time to attempt to discover additional elements of structure in biological materials when the structural details are clearly beyond the theoretical capabilities of the optical instruments employed. Nevertheless experimental efforts to do so continue to be made and to be reported from time to time by workers who are not aware of the limitations or who believe that structural details can be demonstrated in violation of the theory. Such purported structural elements invariably turn out to be diffraction patterns or related artifacts. How much more profitable if the same time were spent in other directions, perhaps with an electron microscope, in which case the theoretical resolving power is several hundred times greater.

There is admittedly some danger in following too slavishly boundary conditions that may be imposed by theory, particularly where theories have not been well established. In research, it must be recognized that "fences are built for those who cannot fly." At the same time one should avoid wasting energy needlessly by flying against an impenetrable barrier.

Knowledge of the theoretical and experimental limitations to the accuracy of a research method or technique has not always been readily accessible to the laboratory worker. Frequently the boundaries have been recognized only after the expenditure of valuable time in a fruitless experiment. Hence, one of the chief aims of this book has been the presentation of information with respect to the limitations of the methods discussed in the various chapters.

2. Instrumentation and Technique

In discussing the research applications of electromagnetic theory, Ernest O. Lawrence used to take great delight in pointing out what could be done if one possessed infinite experimental skill. While conceding that some laboratory scientists are most ingenious, I doubt if any would boast of having infinite experimental skill. Consequently it is usually essential to take into consideration the experimental limitations of a research method. Since barriers to experimental progress today may not exist tomorrow, a recurring evaluation of the limitations is indicated.

Great strides in instrumentation have occurred in recent years, even though the biologist's prayer for a twin-dial light source "box," with one dial for wavelength and the other for intensity, has not yet been answered. Waves of progress in research frequently follow the

introduction of a new tool or a markedly improved instrument. Therefore great efforts toward the elimination of experimental limitations in research seem highly justified.

Where improvements in instrumentation must be of a radical nature in order to secure a really significant advantage, it may not be worth while in a given research program to struggle for a minor gain. For example, the increase in *resolving power* of the optical microscope, which was obtained at great cost and inconvenience by developing quartz lenses for the ultraviolet region, represented a predictable but relatively minor gain (although it is a valuable instrument for other reasons). The *radical* improvement in resolution, amounting to several orders of magnitude, was found three decades later in a completely different type of instrument, the electron microscope.

The accuracy of instrument calibration in critical experiments must be checked by the experimenter. The purchase of a high quality instrument from a manufacturer of integrity does not relieve an investigator of the responsibility of verifying the correctness of its calibration. The calibration may be a function of time, position, temperature, or abuse. Nominal values are often given, as for the magnifying powers of microscopic objectives and eyepieces, it being assumed that the user will make an adequate calibration under the exact conditions employed. For many types of laboratory apparatus, it is possible to have calibrations certified by the National Bureau of Standards. In any event, the question of calibration accuracy must be considered in all careful experiments.

Since most research is done in locations where representatives from the various branches of science can be sought for expert guidance without undue inconvenience, it is regrettable that greater advantage is not taken of the opportunity. By consulting more freely with colleagues about the implications and the possibilities of research problems, errors could be avoided and the general quality of the investigations improved. Certainly in an age of specialization, one should have no feeling of inferiority or embarrassment in seeking the help of specialists. When interested in highly specialized problems, however, an investigator may well discover that the specialists that he needs to consult are not available in his own institution. This situation is especially liable to arise where the experimental limitations of a method are determined by the intricate details associated with recent laboratory instruments and their critical adjustment. In that event, research may often be expedited by a personal visit to

another laboratory. Capable administrators should be able to provide funds for this purpose, basing their decision on the value of the trip in forwarding the worker's research program.

E. CHOOSE BIOLOGICAL MATERIAL CRITICALLY

1. Select Best Genus and Species of Organism

The relative success of numerous investigations depends in large measure on a fortunate choice of biological material. As an example, one may cite the genetic studies with the fruit fly, *Drosophila*. Whether the original choice of *Drosophila* for this important work was primarily accidental or based on careful considerations of its important attributes, I do not know, but the choice has proved to be eminently successful. In addition, an element of luck has since favored its selection in that one of the most important characteristics of *Drosophila* from the cytogenetic standpoint, the existence of the giant salivary gland chromosomes, was not brought to the attention of geneticists until comparatively recently.

While it is clear that certain organisms would not be convenient for genetic experimentation, as for example elephants whether white, gray, or pink, it must be noted that the ultimate qualities of an organism which may possess decisive significance for a given type of investigation are often revealed only after careful study. This may justify a preliminary survey of some duration to discover the best biological material before embarking on an extensive research program. For individuals relatively unfamiliar with experimental organisms, as perhaps physicists who may be working in this borderline field, consultation with several competent experimental biologists is recommended. Since it is not uncommon for a biologist to spend all of his scientific career in the detailed study of a single organism or of a group of closely related organisms, he is in a position to offer invaluable guidance.

The choice of an organism may be determined largely by economic factors. Certainly such factors have dictated the scientific study of numerous cultivated plants and the domestic animals. The widespread use of small laboratory animals for nutritional studies has been based in large measure on size considerations for economic reasons, but their similarity to man in nutritional response has also been a significant item. Other important considerations are: the

length of the life cycle, relative freedom from infection, and favorable
development and behavior in captivity.

Other factors being equal, the most favorable organism from the
standpoint of the particular type of experimentation should be
selected. Since other things are usually not equal, the occasion is
rare when the choice of an experimental organism should be based
solely on its favorable characteristics or response for one specialized
type of investigation. Some of the decisive contributory factors will
be discussed in succeeding paragraphs.

2. Select a Widely Used Organism

Of the large amount of biological experimentation, probably more
than 99% is done on less than 1% of the known species of organisms.
Quantitative studies, whose significance depends largely on the
critical use of special physical methods, are conducted usually on a
few standard biological materials. Most genetic studies involving
physical agents for the production of mutations, for example, are made
on *Drosophila melanogaster* and on *Zea mays*. Nutritional studies are
confined largely to white rats, although use is made also of white mice,
rabbits, guinea pigs, and other domestic animals. Monkeys and
dogs are less frequently objects of study, but are important in that
their response often closely parallels that of man. A list of the bio-
logical materials most commonly employed for quantitative investiga-
tions of various types would be surprisingly brief.

In initial or exploratory studies of some supposedly new phenom-
enon, it may be thought that the choice of biological material is
relatively immaterial as far as the outcome of the experiment is con-
cerned. In that event, it would seem most logical to select a very
widely used organism. The latter course probably would be advisable
even if some slight experimental disadvantage were incurred. As a
consequence, the information obtained can be readily evaluated by
any one of a large group of experimenters, enabling them to make
direct comparisons with other reactions of the same organisms. This
course also favors the wider dissemination of the results of research by
increasing the potential audience to include those working with the
same organism on other problems even though not closely related in
nature. The increased value which attaches to an experiment owing
to such intercomparisons will probably exceed any apparent tem-
porary advantage that would be gained by the casual employment of a
relatively unknown genus or species of organism. As an example,

Auerbach, in investigating the production of chemical mutations by the use of mustard gas, did not select a genetically unknown organism, perhaps the tsetse fly, but chose the best known and most widely used genetic material—*Drosophila*. His results were immediately comparable to similar effects produced by X rays and other mutagenic agents.

There is a rather obvious danger inherent in any unanimous choice of a single organism for a given type of investigation. Important discoveries are likely to be overlooked. For example, the guinea pig, for which vitamin C is essential, stands almost alone among mammals with respect to this nutritional characteristic. The discovery of the vitamin nature of nicotinamide probably was long delayed because rats only were used as the experimental material and most strains of rats do not require it as a dietary ingredient. The later use of dogs and monkeys led to the discovery.

To avoid the systematic errors which may arise from the use of a single biological material, diverse organisms must eventually be employed. Since many laboratories are not adequately equipped to provide such broadened facilities, perhaps this later phase may be done more conveniently elsewhere by some other worker. Many confirmatory experiments are of this type.

3. Select a Genetically Constant Organism

The poet that rhymed: "A primrose by the river's brim, A yellow primrose was to him, That and nothing more" was certainly not deploring a lack of knowledge concerning the relative abundance of its dominant versus recessive genes or to what extent the primrose was homozygous for either. Much of the older biological literature possesses dubious quantitative value owing to the fact that inadequate control was maintained over the genetic constitution of the experimental material. Workers in experimental biology today, however, must become increasingly cognizant of the laboratory importance of using organisms with known genetic constitutions. This point should be stressed all the more in a book of this character since the physical methods discussed may be employed by individuals not fully conversant with the biological variables.

This genetic problem is relatively simple, though not always easy, in the case of bacteria and other unicellular organisms since cultures may readily be obtained from a single parent. In this manner one can be assured that all of the organisms possess a common heritage. Even

here it is necessary to avoid species or strains known to be highly
variable in their characteristics since this variability is probably best
explained at present on the basis of easily mutating genes.

With higher organisms it is much more difficult, if not impossible,
to secure a pure line. Much progress has been made in the case of
standard laboratory animals in that they have been inbred over a few
hundred generations. Likewise some plant materials are rather well
known genetically. Even though vigor and disease resistance in these
hgihly inbred strains may be unfavorable, still their use is indicated
in numerous experiments where otherwise the detection and the inter-
pretation of the experimental effects might be jeopardized or com-
promised.

To employ a single inbred strain to study an experimental effect is
not sufficient, however, if generalized conclusions are to be reached.
It is then necessary to use several inbred strains which span the type
of reaction or effect under investigation. This is analogous to the
situations, mentioned in the previous section, in which the use of
diverse biological materials are indicated.

Several unique solutions of the genetic problem have been found
for higher plant and animal experimentation. One of these is the
widespread use of tissue cultures. Although the tissue culture method
was developed for its superior advantages with respect to environ-
mental control, the elimination of the genetic constitution factor is
clearly of great importance. Not only numerous types of animal
tissues can be cultured almost indefinitely, but also various plant
tissues. A second solution is available for those plants which may be
propagated by cuttings, tubers, grafting, and related methods. With-
out passing through a sexual cycle, they thus achieve a constancy in
their genetic constitution. The possibility of eventual changes is not
completely eliminated, however, since random mutations may occur
in somatic tissue.

F. CONSIDER TIME A FACTOR

Once upon a time, when research progressed at a much slower
pace, a scientist could attack a problem in a leisurely fashion without
being disturbed unduly by the thought that another investigator
might find the answer first. At a somewhat later date in the history of
science it was customary for individuals to stake out claims and
reserve to themselves for rather indefinite periods of time the privilege
of conducting investigations on certain limited aspects of a subject.

This was relatively effective in avoiding duplication of effort in some fields, but it is not a very practicable procedure today, although the custom still exists to some extent on a courtesy basis. As a consequence of the competitive activity in modern research, the time factor must be seriously considered when designing an experimental program. Otherwise results may be obtained only after the same experiment has been performed by someone else. Certainly it can be said of a large percentage of all research problems that if they were not investigated by one group of individuals, they would be studied by another group, and within a comparatively short time. In fact it is not at all unusual for practically simultaneous publication to occur independently in different parts of the world on essentially the same research problem.

1. Time and Equipment

Time and equipment may interact in several ways to determine the fruitfulness of an experiment. The use of obsolete apparatus, for example, may result in the production of obsolete data. This would be an obvious waste of time. Equipment may become obsolete because it is not sufficiently accurate, lacks desirable resolution, or it may be inefficient and time-consuming compared to modern apparatus.

Since time is one of the most valuable assets of research workers, some effort may well be expended on methods of conserving it. The use of recording equipment might be mentioned in this regard. On the other hand, it is not worth while to develop automatic instrumentation for just a few measurements. The purchase of commercial apparatus from which the sources of error and inefficiency have been removed (in other words, the "bugs" have been eliminated) usually means an appreciable saving in research time. It is to be regretted that much high quality research talent has had to be devoted in the past to the development of often low quality apparatus. This false economy, which in effect robs science of many fruitful experiments, is an indication of poor administration.

Many research problems cannot be attacked to advantage because instrumentation is still inadequate. It is therefore debatable whether an investigator not well qualified in instrumental development should undertake such problems. It is quite likely that his efforts might be more fruitfully spent in other directions, especially when indications already exist that satisfactory apparatus may become available in the foreseeable future.

The advisability of inaugurating co-operative research programs should at times be given serious consideration. There are occasions when research problems, or perhaps only limited phases of an investigation, can be studied to great advantage at another institution with equipment already in operation. It is regrettable that this is not done more frequently and that institutional administrations do not have provisions that would readily facilitate co-operative research arrangements. Postdoctoral fellowships have done much to realize this objective for a selected group.

The various marine biological laboratories have been successful in providing research facilities that have been widely used for investigations during the summer months, the expenses often being defrayed in part by the worker's home institution. Not only specialized equipment, perhaps unavailable to the investigator at home, but also specialized biological materials that can be studied to advantage only at a marine location, combine to make this a very satisfactory arrangement.

The current trend toward more expensive and elaborate apparatus is making it correspondingly more difficult for scientists in some institutions to explore the research fields of their greatest interest. In some cases, financial arrangements could be made with the service laboratories of a neighboring university or other scientific institution in order to overcome this handicap. With modern communication facilities it should be readily possible, for example, to have specimens for electron microscope observation prepared at one institution and the photographs taken at another. The same can certainly be said for analyses of stable isotope samples by the mass spectrometer. In fact, many types of analyses might well be made by commercial laboratories at a financial saving to the sponsoring research institution, when all the hidden charges are considered.

2. Time and Personnel

Just as the time factor may have an important bearing on the outline of an experimental program from the standpoint of the equipment available for expediting research, so it likewise has an influence relative to the question of personnel. Some research problems can be met efficiently only by relatively large research groups. Individuals should hesitate to undertake problems by themselves where it is clear that progress can be at best extremely slow, especially if the risk is great that the same problems may be concluded first by a group that is

better prepared to expedite them. It seems increasingly likely that
future research will require more co-operation between individual
scientists rather than less.

Some experiments by their very nature impose demands on time
which exceed the human life span. Where investigations of this kind
must end with the life of the experimenter, the scientific returns may
not be commensurate with the investment in time and energy. It is
therefore partly meaningless to inaugurate such long range experi-
mental programs unless facilities can be assured in advance for their
continuance into the indefinite future. Examples of this type of in-
vestigation occur in the field of experimental genetics where life cycles
are long, as for elephants, turtles, and certain species of trees. In the
same classification are many of the observational problems in the
field of human genetics, which suffers under the great handicap of not
being an experimental science. Since it is conceivable that it may
become an experimental science in the future, perhaps many problems
in the field should await that time.

G. SATISFY IMPORTANT TECHNICAL DEMANDS

1. Design Experiments

In the words of Fisher, who has written an excellent book on the
design of experiments (7): "It is possible, and indeed it is all too fre-
quent, for an experiment to be so conducted that no valid estimate of
error is available. In such a case the experiment cannot be said,
strictly, to be capable of proving anything. Perhaps it should not, in
this case, be called an *experiment* at all, but be added merely to the
body of *experience* on which, for lack of anything better, we may have
to base our opinions."

To make the best use of statistical methods, it is not sufficient to
attempt an analysis subsequent to the recording of experimental ob-
servations, but the experiment must be planned from the beginning
with the demands of the statistical analysis in mind. As a conse-
quence numerous economies often can be effected in the conduct of
experiments without an accompanying decrease in the reliability or
the significance of the results. An example of this type of application
is the determination of the most efficient number of organisms to
serve as controls and for the various parts of an experiment in order to
achieve comparable accuracy throughout. Moreover, it is possible
that a careful analysis of the schedule for an investigation with respect

to statistical factors may indicate in advance that the execution of the
program as originally planned is not likely to yield meaningful results.
Assistance in designing experiments may be obtained from the several
books listed in the bibliography (*7–11*) and from professional stat-
isticians.

Experimental work should be planned on a scale adequate to yield
unambiguous results.   Halfway attempts are uneconomical, often
quite futile, and may be misleading.   In exploratory investigations,
the planning usually is not elaborate, but even preliminary experi-
ments should be conducted in an efficient manner.

2. Control Environmental Factors

Adequate quantitative control of the various environmental con-
ditions constitutes the most difficult part of many biological experi-
ments.   While it may be relatively simple to hold one or two condi-
tions constant at any one time, the problem rapidly grows in com-
plexity when several factors must be controlled simultaneously.
Since the potentially variable factors are numerous, it is rarely pos-
sible to regulate all of them in a completely satisfactory manner.   The
alternative procedure of employing a statistical method of control,
discussed in some detail by Fisher (*7*), should not be overlooked; it
may enable a simultaneous evaluation not only of the various en-
vironmental factors separately but also of their mutual interactions.

Since it is almost never possible to control all the physical and
chemical factors to the extent that one would like, it may prove de-
sirable to select and control the particular environmental influences
likely to exert the greatest effect on the results of a given investigation.
Examples of physical factors could include such items as tempera-
ture, humidity, light intensity, electrical and acoustic fields, static
pressure, and perhaps still others.   With regard to chemical influences,
it will probably be necessary to control the chemical composition of
the medium including its hydrogen ion concentration.   Since it is
well known that biological systems may be very sensitive to exceed-
ingly small quantities of essential elements or compounds, the control
of the chemical environment often constitutes a limiting factor in an
investigation.

Great care is required to insure the point-to-point uniformity of a
particular environmental condition throughout a system containing
biological materials.   The mere installation of a humidistatically
regulated container, for example, does not necessarily guarantee that

the humidity remains constant at each point in the system. A similar difficulty is met with regard to the uniformity of any of the several factors that may be subject to regulation, whether it be the concentration of the chemical constitutents of the medium or the absorption of radiant energy at various depths in a tissue.

In experimental research, where an attempt is made to achieve control of numerous variable factors, it would appear only logical to make the observations objectively wherever possible. In fact, this has been done rather widely. For certain classes of observations this striving for objectivity can be pushed to the point of diminishing return. It would be quite uneconomical, for instance, to insist that all data be recorded automatically.

Photographic observations are quite readily made and consequently have played a valuable role as scientific evidence. In the use of the biophysical methods discussed in this volume, however, the photographic evidence itself will usually require an objective quantitative analysis. A case in point is the densitometric analysis of photographic negatives for various purposes. Although comparisons are often made subjectively, a distinct advantage is gained by using an objective type of microphotometer. Automatic recording equipment is becoming increasingly available so that such variables as temperature, hydrogen ion concentration, and atmospheric gas composition can be continuously registered during an experiment. The same is true of a number of electrical quantities. Especially valuable in this connection is the use of cathode ray oscillographs for recording transient phenomena. For light measurements, the use of a thermopile or photocell together with a recording galvanometer has made it easy to eliminate the subjective factor. Motion picture photography is becoming increasingly useful.

The weaknesses of photoelectric and photographic methods of recording information must not be glossed over lightly, however, They can yield biased data. Insufficient attention has been given to their shortcomings on the part of some biologists in the past. The color sensitivity of a photocell may differ markedly in wavelength range and in relative value from that of the human eye, for example; and books have been written to explain what every scientist needs to know about the behavior of photographic emulsions.

3. Employ Standard Units of Measurement

During the exploratory stage of research in a new field, it often happens from necessity that many measurements are made which are

based on makeshift or arbitrary units. As an example, the earlier investigators of the biological effects of X radiation could not measure the radiation dose in unique fundamental terms but had to resort to lengthy statements of conditions whose value in turn depended on the particular commercial brand of X-ray tube employed. This unsatisfactory procedure persisted for many years, although it was clear that the available information was insufficient to permit an accurate quantitative analysis of the data or to enable other workers to duplicate the experiments. Quantitative progress was thus curtailed until a standard unit of measurement was adopted. For current research to be published in this field without relating the measurements in a known way to the roentgen unit, however, would appear to be inexcusable. Considering the great amount of time and effort devoted to the conduct of a scientific investigation, it is certainly foolish not to take the additional steps that may be necessary in order to record the measurements in a reproducible unit known or available to all. Another pertinent illustration is the measurement of a liquid dose or ration in terms of drops. Although rather accurately reproducible by the original investigator, it becomes quite hopeless for others to duplicate measurements satisfactorily unless they are specified also in some absolute unit.

Many investigations of a biophysical nature require the use of one or more physical or chemical constants for the computation of the data. Therefore, it should be realized that the accepted values of these "constants" are subject to change. The atomic weights of the elements are revised almost annually and other fundamental quantities likewise must be periodically subjected to review. The accepted value of the electronic charge, for example, has changed by nearly three-fourths of one per cent during the past twenty years. Thus the data and computations of numerous studies which have been based on the older values, should be altered by a similar amount. An investigator not only must be alert to the best values of the moment for a given constant but also must consider how the revised values will influence the results of earlier investigations of similar problems with which he may wish to make comparisons.

Since present experimental research is very heavily indebted to the existing scientific foundations built by others, it is obligatory on all to make available the results of their work in an intelligible form. Among other things, this demands a statement of the numerical values of the physical constants that have entered into the calculations.

Without this, future workers will not be able to use the data in a strictly quantitative sense.

4. Use Appropriate Degree of Accuracy

In almost any experiment, some of the factors usually can be measured with a high degree of precision, others with an entirely satisfactory accuracy, while perhaps still others with only a comparatively crude subjective estimate. Since the various factors should all be measured with about the same order of accuracy, it is often somewhat ridiculous to go to great lengths to measure one item very precisely while other items in the same experiment are little more than guesses. If the same relative precision is to characterize all aspects of a particular investigation, an increase in the accuracy of the least precise items in an experiment should be the first point of attack in any effort to improve the over-all accuracy.

Although quantitative measurements constitute the basis of experimental science, the necessity for accuracy may be overemphasized. The development of quantitative spectrographic analysis, for example, probably was delayed many years because physicists stressed the possible sources of inaccuracy in the method; however, they did not recognize that it had other tremendous advantages over existing methods and that the accuracy available was sufficient for numerous applications. Today a useful technique of quantitative analysis even employs a modified Bunsen burner as a source.

Extreme accuracy can often be obtained if one is willing to pay an extreme price. Unless the extra effort can be justified in terms of relative usefulness of the data either in the same or other experiments, one should not go to extremes. For example, one might embark on a program to determine the melting points of all organic compounds to a precision of a hundredth of a degree, but would this serve any justifiable purpose either theoretically or experimentally in the foreseeable future?

II. ANALYZE DATA OBJECTIVELY

Experimental data should usually have the benefit of analysis by statistical methods. By definition these methods are especially adapted to the elucidation of quantitative data affected by a multiplicity of causes. Since much biological experimentation yields data wherein the organisms or systems studied are influenced by

numerous causes, both known and unknown, it is almost essential to resort to a statistical analysis. That proper evaluation of the data is not always made has been shown quite strikingly in a survey by Dunn (*12*), who analyzed 200 articles in the field of medical physiology, selected at random from American journals prior to 1930. Dunn concluded:

"(a) In over 90 per cent statistical methods were necessary and not used.

"(b) In about 85 per cent considerable force could have been added to the argument if the probable error concept had been employed in one form or another.

"(c) In almost 40 per cent conclusions were made which could not have been proved without setting up some adequate statistical control.

"(d) About half of the papers should never have been published as they stood: either because the numbers of observations were insufficient to prove the conclusions or because more statistical analysis was essential."

This disgraceful indictment refers to a condition which existed twenty years ago, but there is still reason to stress the value of a statistical analysis of research data even though scientific workers generally have become increasingly statistics-conscious in the past two decades. If such a waste of research time and money as indicated in the study by Dr. Dunn were to continue, it might ultimately lead administrators to impose regulations that would not only curtail freedom of investigation but also stifle research initiative.

Statistical methods furnish a valuable tool in the conduct of experimental research for a number of reasons. Complicated sets of observations often can be described briefly and in simple terms. The use of an objective criterion of accuracy furthers a critical attitude toward experimental techniques and serves as a check on their reliability. Distinctions between significant and nonsignificant experimental differences can be established in terms of an objective technical index. Thus meaningless and unwarranted conclusions can often be avoided, saving both embarrassment and the time of other workers in fruitless attempts to confirm the results. Not the least important benefit from a careful statistical analysis of research data is the confidence created in the results on the part of the original investigator.

Statistical methods, as valuable as they are conceded to be, are not a panacea for all the ills that befall experimentation. Their use demands that the original data be of high quality. The observations must be made meticulously and recorded accurately, with neither sins of omission nor commission. It may be not only valueless but

also quite misleading to treat poor observations statistically. Infinite experimental skill would make statistical analysis unnecessary, but even infinite statistical skill would possess little value if only bad experimental observations were available.

Errors, by their very nature uncertain things, may be classified into several categories. The expression "standard error" refers to a mathematical concept employed as a measure of variation in connection with the Gaussian error curve. It is applied to that class of errors that are presumed small and accidental in nature, and that possess a given symmetrical distribution about a mean value. The usual statistical formulas are strictly applicable to this class of errors only. Other categories are known as constant errors, systematic errors, and blunders. An example will bring out the distinction. In Millikan's measurements on the electron by the oil drop method, the value of the charge seemed to change with the size of the oil drop. This apparent variation was discovered later to be due to a "systematic" error, which was introduced into the calculations through the use of an inadequate equation for the rate of fall of tiny spherical bodies in air; this error was eliminated before Millikan published his final results in 1917. However, it was not discovered until 1928, when Birge carried out a careful recomputation from the original data, that the published value for the electronic charge was in error because its initial calculation involved an obsolete value for the velocity of light and, further, was based on the mistaken assumption that the international volt is identical with the absolute volt. These two "constant" errors amounted to −0.004 and were nearly as large as the standard error. (The "probable error" was stated in this experiment as ±0.0038.) Ten years later (1938–39) improved viscosity determinations for air were found to exceed the best value available to Millikan by more than one-half of one per cent, thus revealing another "constant" error. Although the most acceptable value for the electronic charge has changed through the years, it should be noted that the experimental data with the original estimate of statistical error retain their worth in spite of the large corrections made from time to time as "constant" errors have been discovered.

Large errors, or shall we say blunders, come in still another category. Frequently they occur during the recording of observations or they may appear in publications as typographical errors. Statistical criteria do not apply, obviously, but sometimes an error of this type may be revealed during mathematical analysis of the data.

Often in routine determinations, as for example a quantitative chemical analysis, it is customary to run merely a duplicate set of observations. The variables are fewer in this case and more nearly under control. The purpose of the duplication is primarily to avoid gross mistakes or blunders and not to attain that increase in the reliability of the observations that would accrue from a decrease in the standard error. Where the time consumed in repeating a set of measurements may be unreasonably long owing to the complexity of an experiment, it may not be worth while to proceed beyond a single repetition. More can be gained oftentimes by taking increased precautions to eliminate possible sources of systematic and constant errors and by exercising meticulous care in making and recording the observations. The records usually should be complete even though the observations may not be fully understood at the time. The importance of scheduling critical check experiments as a safeguard against unwitting blunders probably cannot be stressed too highly.

I. SECURE EFFECTIVE PUBLICATION

To be meaningful, according to the viewpoint adopted for this chapter, experimental research should contribute factual information that will be of value in correlating, explaining, or predicting behavior. But "facts," in the words of Midgley (21), "are still, and probably always will be, determined by vote." No matter how completely convinced a research worker himself is that he has found a new fact, it does not become generally accepted as such unless and until it receives the favorable vote of a "scientific jury." The more complete and the better established the experimental evidence, the better are the chances of having the results of research readily accepted, but this in itself is not sufficient. It must be considered also an essential duty, inescapably obligatory on the part of the research scientist, to present the evidence forcefully to the "jury."

It is helpful to write in a language that can be read with ease by the majority of the specialists in a field. For this reason chemical research was published primarily in German during the past century. Chemists who felt they had made an important contribution which might be concealed by their native language, perhaps Russian or Swedish, would not only write their articles in German but also publish them in a German periodical. It is entirely possible that a scientific article published in Chinese might well be lost for a number of

years as far as the main body of scientific workers was concerned, although abstracting journals have done much to avoid such occurrences.   An illustration of this language difficulty has occurred in connection with the important discovery of the chromatographic method by the talented Russian botanist Tswett.   Although Tswett realized the importance of his discovery and even published a comprehensive book on the subject in 1910, this new experimental technique remained practically unnoticed until two decades later—the
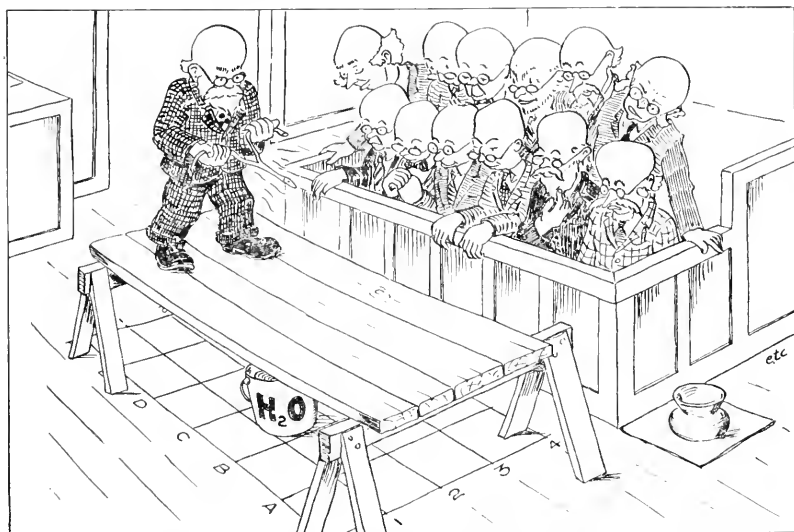
Fig. 3.   "Facts are still, and probably always will be, determined by vote"—Midgely (*21*).   (Courtesy *E. T. Churchman*.)

book had been published only in Russian.   The first English language book on chromatography, a translation of a revised German treatise, did not appear until 1941 (*24*).

The unfortunate selection of a title can be effective in preventing the wide reading of a scientific article.   This seems to have been true of the early articles on the conservation of energy by Joule and Mayer (*26*).   An unsatisfactory title can result in improper indexing and perhaps may lead to its exclusion from the appropriate abstracting journal.   Although titles usually must be brief, and cannot possibly convey the entire content of most articles, some thought should be given to their creation in order that they may serve their purpose in a reasonably effective manner.

Just as it is advisable to write in a widely used language, so it is desirable that the articles appear in journals where they will be noticed by interested scientists.   The historic work of Gregor Mendel, now considered the foundation of modern genetics, was effectively lost to the scientific world for over thirty years presumably because it was published in an obscure journal.   When Mendel's paper was resurrected finally in 1900, it was found desirable to reprint it in a journal of wider circulation.   The reader may wish to speculate on how far the cause of science was retarded by the failure of Charles Darwin to read Mendel's paper, although the paper had appeared sixteen years before Darwin's death in 1882.   Thus Darwin died without knowing the mechanism by which the variations, which he had studied for nearly a lifetime, are inherited.

In the prolific scientific literature of our day, articles can be effectively lost merely by appearing in an inappropriate journal even though the journal itself is well known in another field.   In particular, it is clearly poor policy to publish scientific research of universal interest in journals possessing only local circulation.   The distribution of reprints to interested individuals is becoming an increasingly difficult and ineffective procedure, and should not be considered a desirable alternative in comparison with initial publication in an appropriate journal.

Theses and dissertations that are printed in limited editions, often for private circulation to a restricted list of institutions, do not provide an adequate form of publication in the experimental sciences. It is customary in this country to extract any original contributions contained in dissertations and publish them separately in scientific periodicals of wide circulation.   In a number of universities, reprints of the articles from the latter type of publication are accepted in lieu of typewritten theses.   The publication of original research in a memorial volume also seems unwise unless the volume constitutes a special issue of an established periodical in the field of interest.

Individual workers naturally vary in their preferences as to the type of journal in which they wish to present their results.   It may be that one worker is interested only in convincing his immediate superior, perhaps an individual without a high degree of scientific competence.   Others are looking beyond the bounds of local or state lines. Preferences often depend on the kind of research and on the nature of the institution where it is conducted.   Generally speaking, however, universal recognition is sought for universal truths.   Although

very rare, cases have arisen in which laboratory workers have pre-
ferred apparently to present their evidence directly to the public by
submitting their experimental results initially to the newspapers.
This is frowned upon by most scientists and is not recommended, even
though there may have been a few instances in which it was seemingly
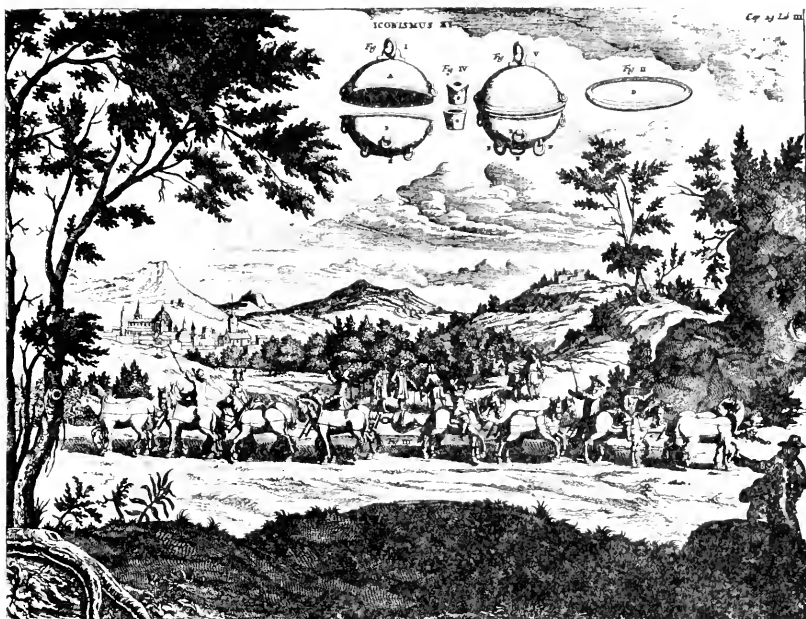necessary to do so owing to the closed minds of those on the "scien-
tific jury."

Fig. 4.   An early historical example of effective publication (after von Guericke).

A conscious attempt to select the most appropriate medium of
publication should always be made.   Whether an article will be ac-
cepted for publication by the journal chosen may depend on subjec-
tive as well as on objective factors.   Therefore, to some extent, an
element of luck or chance is involved.   Most articles are submitted
to one or more independent referees for approval prior to acceptance.
These referees are fellow scientists acting in good faith, one should
assume, and in the interests of science.   Before becoming bitter
about a rejection slip for a manuscript, it is well to consider not only
the probability aspect just mentioned but also the possibility that
perhaps, after all, it was a fruitless experiment from the standpoint of
some of the factors discussed in this chapter.   Remember, too, that
it is a rare manuscript that cannot be improved by constructive
criticism.

Since most scientists are also human, it is to be expected that
the reputation of a scientific worker, just as the character of a witness,
will weight the decision of even a scientific referee.   Where workers
are publishing for the first time and consequently possess only a local
reputation, it is customary for the work to be sponsored by a known
scientist.   The importance of a reputation for publishing reliable
scientific information is difficult to overestimate.   In their enthusiasm
or desire to establish priority, some workers have a tendency to pub-
lish results based on insufficient data.   One need not assume that
such individuals are intentionally dishonest, but merely that they are
lacking in caution, if not in experience.   When an individual once
establishes an unsatisfactory reputation, it may be extremely difficult
for his work to find acceptance even though his subsequent contribu-
tions may be excellent and appear in leading scientific periodicals.

The reputation of a scientific institution may also be considered an
important part of the evidence.   It has been stated, for example, that
scientists of a well known institution regarded any work done elsewhere
as having two strikes against it already.   Although most scientists
would not care to make such statements publicly, it seems evident
that some really pay little attention to scientific publications coming
from certain quarters.   This attitude is regrettable in that competent
scientific workers may well exist in almost any institution.

The question of effective publication concerns also the related
additional elements of time and space.   The frequency and length
of research articles by an individual investigator are often a matter of
comment.   Some thoughts on this subject have been recorded in a
satirical vein by Clark (20), in the interests of a more mature ap-
proach, who contends it does make a difference how thin authors
slice their scientific papers.

In the presumably faster tempo of present day scientific research,
it is rather doubtful that any scientist could withhold publication
until he had amassed such complete evidence as, for example, that of
Roentgen in reporting the discovery of X rays.   Roentgen furnished
such complete proof that nothing fundamentally new regarding the
physical behavior of X rays was discovered for a period of sixteen
years, even though hundreds of papers on X rays were published
during that interval.   Roentgen quickly got a favorable verdict from
his fellow scientists, partly because of the complete and convincing
evidence in the publication itself, but mostly because his results were
readily reproducible within a matter of days in laboratories all over

the world.   On the other hand, some papers are needlessly subdivided into a ridiculous number of smaller units.   Some discretion in this matter seems indicated in the interest of a favorable reputation among fellow workers as well as for the economical and orderly advancement of science.

References

PHILOSOPHICAL ASPECTS OF RESEARCH

1. Bridgman, P. W., *The Logic of Modern Physics.*   Macmillan, New York, 1927.
2. Bridgman, P. W., *Dimensional Analysis.*   Yale Univ. Press, New Haven, Conn., 1922.
3. Northrop, F. S. C., *The Logic of the Sciences and the Humanities.*   Macmillan, New York, 1947.   Quotation courtesy the Macmillan Co.
4. Mees, C. E. K., *The Path of Science.*   Wiley, New York, 1946.
5. Churchman, C. W., *The Theory of Experimental Inference.*   Macmillan, New York, 1948.
6. Poincaré, H., *Foundations of Science.*   Science Press, Lancaster, Pa., 1946.

DESIGN OF EXPERIMENTS AND THE ANALYSIS OF DATA

7. Fisher, R. A., *The Design of Experiments.*   Oliver & Boyd, London, 1942.
8. Fisher, R. A., *Statistical Methods for Research Workers*, 10th ed.   Oliver & Boyd, London, 1946.
9. Worthing, A. G., and J. Geffner, *Treatment of Experimental Data.*   Wiley, New York, 1943.
10. Mather, K., *Statistical Analysis in Biology*, 2nd ed.   Interscience, New York, 1947.
11. Snedecor, G. W., *Statistical Methods, Applied to Experiments in Agriculture and Biology*, 4th ed.   Iowa State College Press, Ames, Iowa, 1946.
12. Dunn, H. L., *Physiol. Revs.*, **9**, 275 (1929).

GUIDES TO THE RESEARCH LITERATURE

13. Crane, E. J., and A. M. Patterson, *A Guide to the Literature of Chemistry.*   Wiley, New York, 1927.
14. Mellon, M. G., *Chemical Publications, Their Nature and Use*, 2nd ed.   McGraw-Hill, New York, 1940.
15. Parke, N. G., III, *Guide to the Literature of Mathematics and Physics.*   McGraw-Hill, New York, 1947.
16. Smith, R. C., *Guide to the Literature of the Zoological Sciences.*   Burgess, Minneapolis, 1942.

17. Soule, B. A. *Library Guide for the Chemist.*   McGraw-Hill, New York, 1938.

18. Kelley, E. C., *Encyclopedia of Medical Sources.* Williams & Wilkins, Baltimore, 1948.

HOW TO WRITE SCIENTIFIC MANUSCRIPTS

19. Trelease, S. F., *The Scientific Paper: How To Prepare It, How to Write It.* Williams & Wilkins, Baltimore, 1947.

20. Clark, W. M., "Evolution toward a Mature Scientific Literature," *J. Bact.*, **27**, 1–18 (1934).

MISCELLANEOUS REFERENCES

21. Midgley, T., Jr., in *The Future of Industrial Research.*   Standard Oil Development Co., New York, 1945, pp. 30–54.

22. Hamilton, T. S., "Some Aspects of Research and the Responsibilities of the Research Worker," an unpublished Sigma Xi lecture delivered at Iowa State College, December 5, 1946.

23. Livingston, B. E., "The Choosing of a Problem for Research in Plant Physiology," *Plant World*, **15**, 73–82 (1912).

24. Zechmeister, L., and L. Cholnoky, *Principles and Practice of Chromatography.* Wiley, New York, 1941.

25. Crane, E. J., *Chem. Eng. News*, **25**, 1363–7 (1947) (nomenclature for D,L-amino acids, etc.).

26. Planck, M., *Naturwissenschaften*, **30**, 285–306 (1942).

OSMOTIC PRESSURE MEASUREMENTS

DAVID R. BRIGGS, *University of Minnesota*

A. GENERAL

1. Definition of Osmotic Pressure

When two phases, one consisting of a pure solvent and the other consisting of a solute dissolved in that solvent, are placed in contact with the opposite sides of a semipermeable membrane, that is, a membrane permeable to the molecules of the solvent but impermeable to the solute molecules, it is observed that a movement of the solvent (osmosis) occurs across the membrane. The direction of movement (if the membrane plays no part other than that of a barrier to the solute) is from the side of the pure solvent toward the side in contact with the solution. The net movement of the solvent molecules can be stopped (*i.e.*, brought to an equilibrium state wherein the passage of solvent molecules in each direction across the membrane is equal)

by the application of mechanical pressure to the solution side of the membrane. This mechanical pressure, which will be exactly equal to and opposite in sign to the difference in the diffusion pressure, or escaping tendency, of the solvent in the two phases, is called the osmotic pressure of the solution. It is a measure of the extent to which the thermodynamic activity of the solvent molecules has been reduced by the presence of the solute molecules in the solution. The osmotic pressure of a solution may be defined as the mechanical pressure that must be applied to the solution in order to bring the solvent in the solution to the same escaping tendency, chemical potential, or partial molar free energy value characteristic of the pure solvent at the same temperature.

The osmotic pressure of a solution is one of the so-called colligative properties of the solution, others being the lowering of the freezing point, the elevation of the boiling point, and the lowering of the vapor pressure. These properties are all reflections of the lowering of the thermodynamic activity of the solvent by the solute and are, in dilute ideal solutions, proportional in degree to the number of particles, ions, molecules, etc. of solute present per unit number of moles of solution. When the solute concentration is sufficiently dilute and the entire effect of the solute upon the activity of solvent is due to an ideal entropy of mixing, and if the effect upon any one of the colligative properties can be measured for a solution of known weight composition, the molecular weight of the solvent being known, the number average molecular or particle weight of the solute can be determined. In any case in which the effect upon any one of the colligative properties of the solution can be measured, the corresponding effect upon any other colligative property can be calculated.

2. Origin of Osmotic Pressure

Many attempts have been made to arrive at a kinetic explanation of osmotic pressure and the other colligative properties of solutions, but none has proved entirely satisfactory in a quantitative sense. Van't Hoff observed the analogy between the laws describing osmotic pressure versus solute concentration and those describing gas pressure versus gas concentration and conceived the idea that osmotic pressure results from a bombardment of the membrane by the solute molecules in a fashion similar to the pressure of a gas arising from the kinetic bombardment of the walls of the containing vessel by the gas mole-

cules.   The fact that the postulated pressure of the solute molecules
is restricted in its manifestation to the region occupied by the solvent
(*i.e.*, no such pressures are actually existent upon the walls of a vessel
containing the solution) has tended to work against the acceptance of
the completeness of the analogy. Others have pictured osmotic
pressure as resulting entirely from solvent bombardment, the number
of solvent molecules reaching the side of the membrane in contact
with the pure solvent beng pictured as greater than that from the
solution side because of the cross-sectional area occupied by solute
molecules on the solution face of the membrane. However, it is
pointed out that the effect of the solute is not proportional to the size
of the solute molecule but only (in dilute solutions) to the number
present per unit volume of the solution. If it is assumed, however,
that the solute in some way causes a decrease in the solvent pressure
in the solution and that this decrease is proportional to the number of
solute particles present per unit of solvent molecules, we arrive at a
basis for the thermodynamic treatment of osmotic pressure together
with that of the other colligative properties of the solution in which
these effects are related quantitatively to the escaping tendency or
fugacity of the solvent in the solution compared to that of the solvent
in the pure state.   There is still to be answered the question as to the
actual kinetic mechanism by which this decrease in solvent fugacity is
brought about.   In the absence of a satisfactory kinetic picture it is
necessary only to bear in mind that, for some reason connected with
the presence of solute molecules, the partial free energy, escaping
tendency, or activity of a molecule of solvent is lower in the solution
than in the pure liquid, and that the transfer of solvent from pure
liquid to the solution, whether through a membrane semipermeable
to the solute or through the vapor phase (which also may be regarded
as a medium permeable to solvent molecules but impermeable to
solute), will occur with a loss of free energy and will, therefore, be a
spontaneous process.   External pressure applied to the solution will
cause an increase in the vapor pressure of the solvent in the solution,
i.e., increase its escaping tendency, activity, or partial free energy.
Where a membrane separates the 2 phases and when this pressure is
just great enough to bring the activity of the solvent in the solution
to that of the pure solvent, no further net transfer of solvent will
occur across the membrane and the applied pressure will be a measure
of the osmotic pressure of the solution.

3. Importance of Osmotic Pressure Determinations

The determination of the osmotic pressure of solutions has two realms of primary usefulness in biophysics.   The first is in plant and animal physiology, where the osmotic pressure is of chief interest and importance because of its capacity to influence the distribution of fluids and solutes, to which the membrane may be permeable, across semipermeable membranes in the organism.   Where isolated cells or tissues are being subjected to studies while immersed in solutions, it is usually essential that the surrounding solution be of the same osmotic characteristics as the fluids that bathe them in their normal habitat. When fluids are to be injected into the blood stream, intercellular spaces, or into the cells themselves, a minimum of discomfort or shock to the organism will occur when the osmotic properties of the solution approximate those of the fluids in the locus of injection.   It is necessary for the biologist to know, with a fair degree of approximation, the osmotic pressure of the extra- and intracellular fluids of the specimen with which he is working.   Because the body fluids usually contain a variety of osmotically active solutes for which a semipermeable membrane would be difficult to obtain, osmotic pressures are seldom measured directly in such cases, but are calculated from measurement of the freezing point depressions, the vapor pressures, or a complete chemical analysis of the solutions under consideration.

The quantity osmotic pressure is commonly employed in physiology to express numerically the difference in degree to which the activity of the solvent may in any particular case be maintained across cellular or tissue membranes in the body and as a measure of the concentration work (5) required to bring about this difference.   When solutions of unequal composition exist in dynamic equilibrium on the two sides of a cellular or tissue membrane, it can generally be assumed that the membrane has performed work in bringing about and maintaining this difference.   The minimum amount of work required can be calculated as the summation of the concentration works or minimum free energy changes for all the components of the two solutions, and will be given by the expression:

$$W_{\min} = \Sigma n R T \ln (C_B/C_A)$$

where C_B and C_A refer to the concentrations (more accurately, the activities) of any individual component in solution B (taken as the final solution) and solution A, respectively, n is the number of moles of the component transferred from solution A to B, R is the gas con-

stant, T is the absolute temperature, and Σ is a *summation* symbol that indicates that the value of W_{min} refers to the concentration work for all components of the solutions, algebraic signs being taken into consideration. Osmotic work is that part of the total concentration work that involves changes in solvent concentration alone and is related to the observed osmotic pressure difference, π_0, between the solutions by the relationship:

$$\pi_0 V_1 = RT \ln (N_A/N_B)$$

where V_1 is the partial molar volume of the solvent and N_A and N_B refer to the mole fraction of solvent in solutions A and B, respectively. Osmotic pressure, therefore, will give information concerning the concentration work done on the solvent alone by a tissue membrane, but will tell nothing about the concentration work performed by the membrane on the various solutes that may be present. It would be erroneous to assume, in a system containing a tissue membrane across which there exists no osmotic pressure difference, that no concentration work is being performed by the membrane.

The second realm of usefulness of osmotic pressure measurements in biophysics, and one that usually makes use of the direct measurement of the quantity, is in the determination of number average molecular weights of the naturally occurring high molecular substances, such as proteins, gums, etc. With such solutes, because of their relatively huge molecular dimensions, it is quite easy to obtain membranes completely impermeable to these particles while easily permeable to the molecules of a solvent. It is also readily apparent that, because of their high molecular weights, these substances must show only small effects upon the colligative properties of their solutions. An example will illustrate why osmotic pressure measurements are resorted to in preference to those of the other colligative properties. One gram molecular weight of any solute dissolved in 1000 g. of water will, for dilute solutions obeying the ideal solution laws, depress the freezing point of the water 1.86°C., elevate the boiling point 0.54°C., and depress the relative vapor pressure of the water by 0.018. The osmotic pressure for such a solution would amount to 22.4 atmospheres (about 1700 cm. of mercury or 23,150 cm. of water) at 0°C. For a protein of a molecular weight of 45,000 dissolved in water to make an approximately 1% solution (1 g. of protein to 100 g. of water) the freezing point depression would amount to 0.00041 °C., the relative vapor pressure lowering would be only 4 parts in a mil-

lion, while the osmotic pressure at 0°C. would amount to 51.4 cm. of
water. Absolute freezing point depressions can be measured, with
highest refinements of technique, to within an accuracy of perhaps
0.0025°C. The vapor pressure of water at, say, 25°C. cannot be
measured with an accuracy exceeding about 1 part in 10,000. The
osmotic pressure, however, can be estimated to perhaps 0.5–1.0 mm.
of water. It is obvious that with such a protein solution the freezing
point depression or the relative vapor pressure change would be un-
detectable but that, where direct measurement of osmotic pressure
can be accomplished, this method will yield results of a sufficiently
high order of accuracy to be useful for the purpose of calculating the
molecular weight of the solute. (Boiling point elevation measure-
ments, while of about the same degree of accuracy as freezing point
depression measurements, are of little value for biological systems
and substances because of the many changes that may occur in such
systems when heated.) Because of the slight effect produced per
molecule on the relative vapor pressure or freezing point these
methods when applied to the determination of molecular weights of
solutes are most valuable only when the molecular weight of solutes
is less than about 4000.

While the classical works of Morse and Frazer (*32,33*) and of
Berkeley and Hartley (*24*) showed that osmotic pressure measure-
ments of high accuracy can be made on solutions containing low
molecular solutes, they also emphasize the great difficulty encountered
in preparing membranes semipermeable to such solutions. In gen-
eral, it is the conclusion of experimental studies that for solutes of
molecular weights less than about 10,000 to 15,000 it is very difficult
to prepare membranes sufficiently semipermeable to allow for de-
pendable measurements of osmotic pressure. On the other hand,
for a solute of molecular weight 500,000, dependable extrapolation of
data to an intercept value requires measurements at concentrations
for which the observed osmotic pressure is less than 0.5 mm. of water
and values so obtained become inaccurate because the probable error
of the measurement approaches or exceeds the actual measured
value. It is clear, therefore, that the usefulness of direct osmotic
pressure measurements in the determination of molecular weights
will be confined to those solutes having molecular weights between
10,000 and 500,000 and will be of true significance for this purpose
only if the solute is homogeneous. Molecular weights calculated
from osmotic pressure data obtained on a mixture of nondiffusible

solutes will constitute a number average molecular weight of the components of the mixture but will give little information concerning either the range of the molecular weights of the components or the relative amounts of the various components present.   The number average molecular weight obtained from osmotic pressure measurements is nevertheless of considerable value in following the degree of polymerization during the preparation of high polymers.   The values of molecular weights obtained from osmotic pressure measurements will be dependable in many cases only when calculated from extrapolated values of the osmotic pressure/concentration relationship obtained at infinite dilution and when such extraneous effects as may arise from a Donnan membrane equilibrium are avoided or taken into account.

The use of membranes impermeable to high molecular substances but freely permeable to other solute constituents of a solution makes possible the determination of the fractional effect of the higher molecular constituent upon the colligative properties of the solution.   This circumstance is taken advantage of in the estimation of the so-called colloid osmotic pressure (oncotic pressure) of blood or of other biological fluids that commonly contain both high and low molecular solutes, and also constitutes an important reason why direct osmotic pressure measurements of the molecular weights of high molecular substances is to be favored over the other methods mentioned above, *i.e.*, freezing point and vapor pressure methods.   It is often difficult to prepare proteins, gums, etc. entirely free of low molecular impurities.   On a weight basis these low molecular solutes will have a relatively great effect upon the colligative properties of the solution and, even though present in very small amounts by weight, will lead to large errors in the values of the molecular weight of the high molecular component when calculated on the basis of the total effect upon the properties of the solution obtained by freezing point depression or vapor pressure decrease measurements. In the case of osmotic pressure measurements, with membranes such as those referred to, the low molecular substances distribute themselves at equal concentrations on both sides of the membrane; this leads to no contribution by them to the observed osmotic pressure.

B. THEORETICAL CONSIDERATIONS

Before considering methods and their possible applications in biophysics, it seems desirable that a short summary of the theory in-

volved be presented. No attempt will be made to cover, in an adequate manner, the many ramifications of osmotic theory but some mention of the factors that determine the validity of osmotic measurements under various conditions must be considered prior to a discussion of methods. This summary will be limited to a brief exposition of: (1) the osmotic pressure relationships in dilute solutions of low molecular substances where this property and the other colligative properties of the solutions approximately obey the ideal solution laws as set forth by van't Hoff and Raoult, (2) the osmotic pressure relationships in more concentrated solutions and in solutions of high molecular substances, and, finally, (3) the effects of Donnan membrane equilibria upon the osmotic pressure relationships in solutions of colloid electrolytes. In the following theoretical section, subscript 1 refers to the solvent and subscripts 2 or higher refer to solute components of the systems under consideration (*1*).

1. In Dilute "Ideal" Solutions

A satisfactory mathematical treatment of the laws governing osmotic pressure and the other colligative properties of dilute solutions has been derived on the basis of thermodynamic considerations alone. The definition of osmotic pressure given above is arrived at on this basis. When equilibrium has been established in an osmotic cell consisting of a pure solvent phase separated from a solution phase (containing the same solvent) by a semipermeable membrane, an excess pressure equal to the osmotic pressure, π, will prevail upon the solution. The work required under this condition to transfer one mole of the solvent from an infinite volume of the solution into the pure state is equal to $\pi \bar{V}_1$, *i.e.*:

$$\pi \bar{V}_1 = -\Delta \bar{F}_1$$

where $\bar{V}_1$ is the partial molar volume of the solvent and $\Delta \bar{F}_1$ is the work required to accomplish the transfer under reversible conditions and is called the partial molar free energy change of the process. $\Delta \bar{F}_1$ is related to the fugacity of the solvent in the solution, f_1, and to the fugacity of the solvent in the pure state, $f_1{}^\circ$, as well as to the vapor pressure of the solvent in the solution, P_1, to the vapor pressure of the solvent in the pure state, $P_1{}^\circ$, and to the activity of the solvent in the solution, a_1 (referred to the pure solvent as the standard state), by the equation:

$$\Delta \bar{F}_1 = RT \ln (f_1/f_1{}^\circ) = RT \ln (P_1/P_1{}^\circ) = RT \ln a_1$$

The partial molar free energy change for such a process is related also to the partial molar heat content change, $\Delta \bar{H}_1$, and the partial molar entropy change, $\Delta \bar{S}_1$, by the Gibbs-Helmholtz equation:

$$\Delta \bar{F}_1 = \Delta \bar{H}_1 - T \Delta \bar{S}_1 \tag{1}$$

For "perfect" or "ideal" solutions, the mixing of the components (in the liquid state) will have occurred without any deviation from additivity in heat content or in volume, i.e., the heat of dilution is zero and the partial molar volumes will not change with concentration. The total change in free energy will be due to an entropy of mixing that, in cases in which the molecules of the components are of comparable dimensions, will approximate an "ideal" entropy of mixing in that the activity of a component in the mixture will be changed from its activity in the standard state in a manner proportional to its mole fraction.   This relationship is illustrated in Raoult's law, which states that:

$$(P_1{}^\circ - P_1)/P_1{}^\circ = n_2/(n_1 + n_2) = N_2 \text{ or}$$
$$P_1/P_1{}^\circ = n_1/(n_1 + n_2) = N_1 \tag{2}$$

where $P_1{}^\circ$ and P_1 are the vapor pressures of the solvent in the pure liquid and in the solution, respectively, n_1 and n_2 are the moles of solvent and of solute in the solution, respectively, and N_1 and N_2 are the mole fractions of solvent and of solute in the solution.   Under these circumstances:

$$\Delta \bar{F} = RT \ln N_1 = RT \ln a_1 \tag{3}$$

Then, in ideal solutions:

$$\pi \bar{V}_1 = -RT \ln N_1$$

In very dilute solutions, most of which approach "ideality," the partial molar volume, $\bar{V}_1$, approximates closely the actual molar volume, V_1, and we can write:

$$\pi V_1 = -RT \ln N_1 \tag{4}$$

and since $N_1 + N_2 = 1$:

$$\pi V_1 = -RT \ln (1 - N_2)$$

which upon expansion of the logarithm gives the relationship:

$$\pi V_1 = RT(N_2 + \tfrac{1}{2} N_2^2 + \tfrac{1}{3} N_2^3 + \ldots) \tag{5}$$

In dilute solutions, higher terms than the first in N_2 may be ignored and:

$$\pi V_1 = RTN_2$$

Furthermore, at high dilution, N_1 becomes very nearly unity, V_1 can be considered equal to the molar volume of the solution, V, and N_2/V is equal to the concentration, C_2, of the solute in moles per unit volume of the solution. Under these conditions:

$$\pi = RTC_2 = RT(c_2/M_2) \tag{6}$$

where c_2 is the concentration of the solute in grams per unit volume of solution and M_2 is the gram molecular weight of the solute. In this equation, if c_2 is expressed as grams per milliliter of solution, R has the value of 82.07 milliliter atmospheres and π is the osmotic pressure in atmospheres. Equation (6) is a form of the van't Hoff equation and is the limiting or ideal law describing the relationship of the osmotic pressure of a solution to the molar concentration of the solute.

The other colligative properties of a solution can be utilized in the calculation of the osmotic pressure as follows:

From equation (2) it is seen that, in the "ideal" solution, $P_1/P_1{}^\circ = N_1$, from which, by substitution in equation (4):

$$\pi = -\frac{RT}{V_1} \ln \frac{P_1}{P_1{}^\circ} = \frac{RT}{V_1} \ln \frac{P_1{}^\circ}{P_1}$$

Where water is the solvent, at $0\,^\circ C.$, the osmotic pressure, in atmospheres is:

$$\pi = \frac{82.07 \times 273}{18} \ln \frac{P_1{}^\circ}{P_1} = 1245 \times 2.313 \log \frac{P_1{}^\circ}{P_1} = 2807 \log \frac{P_1{}^\circ}{P_1}$$

In a molar solution, the value for $P_1{}^\circ/P_1 = 1.018$, $\log 1.018 = 0.0078$, and:

$$\pi = 2807 \times 0.0078 = 22.4 \text{ atmospheres}$$

In ideal solutions at high dilution it can be shown that the depression of the freezing point, $\Delta T = T_0 - T$, of the solvent in the solution is given by the equation:

$$\Delta T = -(RT_0 T / L_f) \ln N_1 \tag{7}$$

where T_0 and T are the freezing points in degrees absolute of the pure solvent and of the solvent in the solution, respectively, L_f is the molar latent heat of fusion in calories, and R is given in calories per degree per mole and has a numerical value of 1.9864. One calorie is equal to 41.3 milliliter atmospheres and, when L_f is expressed in milliliter atmospheres, R has the same value as in the preceding equations, i.e., 82.07.

From equations (7) and (4) we find:

$$\pi = L_f \Delta T / T_0 V_1$$

When water is the solvent, and L_f is expressed in milliliter atmospheres—$L_f(H_2O) = 59{,}200$ milliliter atmospheres—V_1 is 18 ml. and T_0 is 273 °K. Then:

$$\pi = \frac{59{,}200 \times \Delta T}{273 \times 18} = 12.06 \, \Delta T$$

The molar freezing point depression for water solutions is 1.857 °C. = ΔT. Then:

$$\pi = 12.06 \times 1.857 = 22.4 \text{ atmospheres}$$

2. In Actual Solutions

Many solutions, when sufficiently dilute, obey the van't Hoff law (equation 6) but deviate from this relationship to an increasing degree as the concentration of solute increases. Agreement is commonly found to maintain through a wider change in concentration if c_2 is measured in grams per unit weight of solvent rather than in grams per unit volume of solution.

Deviations from van't Hoff's law in actual solutions may be traced to failure of either of two fundamental assumptions made in the derivation of this relationship. In the first instance, van't Hoff's law may fail, even in "ideal" solutions (where equation 3 holds), because the higher terms in N_2 (equation 5) become increasingly important as concentration of the solute increases. It is repeatedly observed, particularly in studies of the osmotic behavior of high molecular substances in solution, that the rate of increase of osmotic pressure with concentration increases more or less markedly with concentration. Often the excessive increase in the osmotic pressure with increase in concentration is greater than can be accounted for on the basis of

equation (5).   In many cases it is not possible to attain a dilution
sufficient such that the π/c_2 relationship becomes constant, but as has
been shown by Ostwald (6) it is usually possible to obtain osmotic
pressure measurements at concentrations sufficiently dilute so that
when π/c_2 is plotted against c_2 a straight line (constant slope) is ob-
tainable that can be extrapolated to $c_2 = 0$.   It has become common
practice to assume that the value of π/c_2 at $c_2 = 0$ obtainable by this
procedure can be inserted in the van't Hoff equation to yield a de-
pendable value of M_2, i.e.:

$$\lim_{c_2 \to 0} (\pi/c_2) = RT/M_2 \tag{8}$$

This treatment of the experimental results is based upon the assump-
tion that in all such cases the systems are acting as "ideal" as a limit-
ing law, i.e., that $\ln a_1 = \ln N_1$ at infinite dilution.

While in many instances this may approximate the truth, there
is also the probability that in others this assumption will be unten-
able.   Certainly there are many systems in which $\ln a_1 (= \Delta \bar{F}_1)$ is
not definable entirely in terms of an ideal entropy of mixing.   Other
entropy factors may arise from failure of the molecules when in a
solution and surrounded by unlike molecules to retain properties,
such as volume and heat capacity, identical to those exhibited when
they are surrounded by like molecules.   Also the partial molar heat
content term in equation (1) will not be zero in solutions where dilu-
tion is accompanied by evolution or absorption of heat, or by devia-
tion of volume from additivity.   Such changes constitute evidences
of differences in the play of intermolecular forces between similar
and dissimilar molecules.   Of great importance is the likelihood that,
when molecules of very great difference in size and shape are mixed
in solution, the entropy of mixing will not be described by the rela-
tionship, $-\Delta \bar{S}/R = \ln a_1 = \ln N_1$, but by some more complex rela-
tionship.   For example, Flory (7) and Huggins (8) have derived equa-
tions in which, on the basis of statistical considerations wherein flex-
ible long chain polymers are pictured as acting kinetically as seg-
ments rather than as single units per molecule, they describe the ac-
tivity of solvent in a solution in the following terms:

$$\ln a_1 = \ln V_1 + \left(1 - \frac{\bar{V}_1}{\bar{V}_2}\right) V_2 + \mu_1 V_2^2 \tag{9}$$

where V_1 and V_2 are the volume fractions of solvent and solute, re-
spectively, in solution, $\bar{V}_1$ and $\bar{V}_2$ are the partial molar volumes of the

components, μ_1 is a constant characteristic of a given solute–solvent system at a given temperature, and a_1 is the thermodynamic activity of the solvent. The factor, μ_1, which may be considered constant at low polymer concentrations (25 g. per liter or less), appears to depend strongly upon forces acting between solvent and solute molecules, *i.e.*, solvation forces. It is dependent in part upon the heat of mixing ($\Delta \bar{H}$ term) and in part on the departure from perfect randomness of mixing of molecules in the solution ($\Delta \bar{S}$ term other than that of ideal mixing). When $\mu_1 = 0$ and $\bar{V}_1 / \bar{V}_2 = 1$, V_1 is equal to N_1, and equation (9) reduces to equation (3). With flexible long chain molecules as the solute, however, Huggins concludes that equation (9) will better define ln a_1, and by a procedure analogous to that used above in the derivation of the van't Hoff law, he arrives at the following equation as describing the osmotic relationships in solutions containing such molecular species as solute:

$$\frac{\pi}{c_2} = \frac{RT}{M_2} + \frac{RTd_1}{M_1 d_2^2} \; (\tfrac{1}{2} - \mu_1)c_2 + \frac{RTd_1}{3 \; M_1 d_2^3} \; c_2^2 \; + \; \ldots$$

where d_1 and d_2 refer to the densities of solvent and solute, respectively, M_1 to the molecular weight of solvent, and the remaining terms have the same connotation as above. From a plot of the term:

$$\frac{\pi}{c_2} - \frac{RTd_1}{3 \; M_1 d_2^3} \; c_2^2$$

versus c_2, the intercept would yield a value of RT/M_2 (and hence M_2) and μ_1 could be calculated from the slope. Thus:

$$\lim_{c_2 \to 0} \left(\frac{\pi}{c_2} - \frac{RTd_1 c_2^2}{3 \; M_1 d_2^3} \right) = \frac{RT}{M_2} \tag{10}$$

It is obvious that values of M_2 obtained with equations (8) and (10) from a given set of osmotic pressure data will differ to a degree that will be a function of the extent to which the values of $\bar{V}_1 / \bar{V}_2$ differ from unity and the extent to which μ_1 differs from zero.

This short discussion of the osmotic pressure relationships in actual solutions, while necessarily fragmentary, will serve to emphasize some of the limitations of interpretation imposed upon such studies. Thus, while it is possible to calculate the values for the various colligative properties for a given solution from the determined value of any one colligative property (*i.e.*, the value of a_1 determined

by one method of measurement will agree with that determined by any other) it is not possible to predict from such a determination what the value will be at any other concentration than that upon which the measurement was actually made unless the system is either (a) known to obey the laws of ideal dilute solutions or (b) some knowledge is available as to the actual dependence of a_1 upon concentration. For example, the observation that a red blood cell or a muscle tissue slice swells as an "imperfect osmometer" may mean, as is usually assumed to be the case, that the membrane of the cell is not acting in a strictly semipermeable manner to the solutes within and outside the cell; but it should not be forgotten that another possible explanation may be that the cell solute contents do not affect the activity of the solvent in a manner described by the ideal solution laws. This would be the more probable the greater the percentage of the total effect due to high molecular solutes within the cell. Again, where osmotic pressure is used to measure the molecular weight of a solute, particularly where the solute is a high molecular substance, a determination of the osmotic pressure at a single concentration of the solute may be insufficient. Determinations must be made at several concentrations of increasing degrees of dilution such that when a plot of π/c_2 versus c_2 is made, a dependable extrapolation can be attained for the value of π/c_2 at $c_2 = 0$. Even then the calculated value for M_2 obtained by inserting this value of π/c_2 into van't Hoff's equation will be in error if the relationship $\ln a_1 = \ln N_1$ cannot be safely assumed to hold at infinite dilution for the system under study.

3. Donnan Membrane Equilibria

In the foregoing discussion of osmotic pressure theory, it was assumed that the solutes involved were nonionizing substances or, if capable of ionization, that the membrane used was impermeable to all ions so derived, in which case the osmotic pressure would be the summation of the osmotic increments due to each species of particle present in the solution. Many naturally occurring high molecular compounds such as proteins, gums, etc., for which osmotic pressure methods are commonly employed in the determination of their molecular weights, are electrolytes. These substances in solution are capable of ionization, yielding polycharged ions of high molecular weight, to which a membrane may be impermeable while it is completely permeable to the small ions (counter ions).

In an osmotic cell containing such a colloid electrolyte as solute the small "counterions" will tend to reach a condition of equal concentration on either side of the membrane but fail to do so because of the electrostatic attraction for the impermeable component from which they are derived. This situation will, in turn, influence the distribution of other small ions in the system. At equilibrium there will exist a difference in small ion distribution (as electrostatically neutral electrolytes) on the two sides of the membrane that will result in an added increment of osmotic pressure over that due to the nondiffusible (high molecular weight) ions alone.

For illustrative purposes, let us consider a protein that has been titrated with sodium hydroxide to some pH value alkaline to the isoelectric point of the protein (*i.e.*, at such a pH that H^+ and OH^- ions are at insignificant concentrations). The protein salt in solution will ionize to a protein ion bearing z_2 negative charges per molecule and to z_2 Na^+ ions per molecule (z_2 may be called the "valence" of the protein). If this protein salt is dissolved, at a molar concentration C_2, in, say, an aqueous sodium chloride solution of molar concentration C_3, and placed in one compartment (1) of an osmotic cell and the sodium chloride solution (concentration C_3) placed in the other compartment (2) and a pressure is maintained in compartment 1 such that no change in volume of the fluids in the two compartments takes place up to the time when equilibrium is reached, it will be observed that a concentration change will have occurred with respect to the sodium chloride. The concentration of sodium chloride in compartment 2 will be greater than at the beginning of the experiment, while in compartment 1 it will be less than originally. Donnan (*11*) was the first to point out that this phenomenon is the result of the circumstance that the thermodynamic condition for equilibrium in such a system requires that the products of the activities of the diffusible ions be the same on both sides of the membrane, that is:

$$(a_{Na^+})_1 \, (a_{Cl^-})_1 = (a_{Na^+})_2 \, (a_{Cl^-})_2 \tag{11}$$

At equilibrium, assuming the activity coefficients of the ions to be unity, $[Na^+]_1$ will be equal to the sum of that derived from the protein salt and from the sodium chloride present and will therefore (electrical neutrality in each solution being maintained) be greater than the $[Cl^-]_1$. In compartment 2, however, $[Na^+]_2 = [Cl^-]_2$. From equation (11), the relationships $[Na^+]_1 > [Na^+]_2$ and $[Cl^-]_1 < [Cl^-]_2$ must hold at equilibrium so that $[NaCl]_2 > [NaCl]_1$ will also

be true. A concentration increment, x, of sodium chloride will have moved from compartment 1 to compartment 2, as equilibrium is attained. The final concentration of $[Na^+]_2 = [Cl^-]_2$ will be $(C_3 + x)$, while the final concentration of $[Cl^-]_1$ will be $(C_3 - x)$, and of $[Na^+]_1$ will be $(C_3 - x + z_2C_2)$. From equation (11), letting activities equal concentrations, at equilibrium:

$$(C_3 - x + z_2C_2)(C_3 - x) = [C_3 + x]^2$$

and:

$$x = z_2C_2C_3/(4\ C_3 + z_2C_2) \tag{12}$$

The observed osmotic pressure at equilibrium, π_0, must be the sum the osmotic pressure due to the protein ions, π_p, and that due to the unequal distribution of diffusible ions, π_i.

$$\pi_0 = \pi_p + \pi_i = RTC_2 + RT[z_2C_2 + 2(C_3 - x) - 2(C_3 + x)] =$$
$$RTC_2 + RT(z_2C_2 - 4\ x) \tag{13}$$

Combining equations (12) and (13):

$$\pi_0 = RTC_2 + RT\ \frac{(z_2^2C_2^2)}{(4\ C_3 + z_2C_2)} \cong RTC_2 + RT\ \frac{z_2^2C_2^2}{4\ C_3}$$

Since $C_2 = w_2/V_0M_2$, where $w_2 =$ weight of protein in grams, $M_2 =$ molecular weight of protein, and $V_0 =$ volume of 1 kg. of solution when $w_2 = 0$, we can write:

$$\pi_0 = \frac{RT}{V_0}\ \frac{w_2}{M_2}\left(1 + \frac{z_2^2w_2}{4\ C_3M_2V_0}\right) \tag{14}$$

as the expression describing, for *ideal* solutions, the relationships between the osmotic pressure, the valence of the protein (z_2), and the initial concentration of the salt (C_3) (in solutions of uni-univalent diffusible electrolytes). From this expression it is clear that, the lower the value of z_2 and the higher the value of C_3, the less will be the importance of π_i, the osmotic increment due to the unequal distribution of diffusible ions, relative to that due to the protein ion, π_p. For this reason, it is the common practice, when the molecular weight of a protein is to be calculated from osmotic pressure measurements, to conduct the experiment at a pH value close to the isoelectric

point of the protein (value of z_2 near zero) and in solutions of salts of fairly large C_3 values (*e.g.*, 0.1–0.4 M in sodium chloride). Under these conditions the value of the second term in equation (14) should become negligibly small. There are instances, however, in which a protein may become agglomerated or even precipitated when brought to its isoelectric point or when salt concentrations become high. Also the isoelectric point of a protein may shift when salts are added to the solution. In the former case, erroneous and erratic values for π_p would be encountered; in the latter there is lack of certainty as to the pH value at which z_2 is zero.

If z_2 is kept constant (pH constant) and C_3 is kept constant while w_2 is varied, the procedure employed by Scatchard, Batchelder, and Brown (*47*) may be used to obtain the molecular weight of the high molecular constituent. If it is assumed that equation (14) holds as a limiting law (*i.e.*, the protein in solution acts ideally at infinite dilution), when π_0/w_2 versus w_2 is plotted, both the second term (due to π_i) and any nonideal properties of the protein itself in solution will disappear in the limit of infinite dilution ($w_2 = 0$), that is:

$$\lim_{w_2 \to 0} \quad (\pi/w_2) = RT/V_0 M_2$$

and the intercept value of π_0/w_2 will yield acceptable values of M_2.

Deviations of actual solutions from ideality are marked in most instances; a discussion of the relationships involved is too complex to be included here. Reference should be made to papers by Adair (*10,11*) and Scatchard (*9*).

C. METHODS FOR MEASURING OSMOTIC PRESSURE

1. Indirect Methods

When it is desired to know the osmotic pressure of solutions containing solutes of low molecular size for which it is not feasible to prepare membranes to which these solutes are impermeable, or when the solutes, though present mainly as micelles to which the easily obtainable membranes are impermeable, exhibit a tendency to dissociate upon dilution into smaller solute molecules that are diffusible (such as soaps in water solution), it is not possible to employ direct methods for such determinations. Indirect methods, involving the measurement of other colligative properties of the solutions, from which the

osmotic pressure can be calculated, must be resorted to in such instances.   The more commonly measured colligative properties that may be studied for this purpose are the depression of freezing point and the depression of vapor pressure of the solvent in the solutions. As has already been pointed out, these methods are not of sufficient precision to detect the effects of most higher molecular weight substances upon the colligative properties of the solvent with the accuracy required for the calculation of the molecular weights of these substances.   On the other hand, these methods are useful for body tissue fluids where the degree of precision required in determining the absolute effects upon the colligative properties of the solvent may not be so exacting and where, because of the presence of low molecular components, direct measurement of osmotic pressure is not feasible.

Determination of Freezing Point Depression.   For general purposes, a freezing point determination is made by placing the sample, solvent or solution, in a tube surrounded by an air jacket (which acts to retard the heat exchange between bath and sample) immersed in a freezing mixture, the tube being fitted with a thermometer graduated to hundredths of a degree (Beckmann type) and a stirrer.   As the sample cools it is stirred to minimize undercooling.   When the sample is a pure solvent, the degree of undercooling is not important since the liquid remaining in equilibrium with the separated solid is still pure liquid but, when the sample is a solution, the final liquid phase will be a more concentrated solution than before the solid (solvent) phase appears, and the observed freezing point will be lower than the true value (for the original solution) by a factor related to the degree of undercooling, the heat capacity of the solution, and to the heat of fusion of the solvent.   Undercooling must be absolutely avoided or must be corrected for.   It is difficult to avoid undercooling, especially when colloidal components such as proteins, etc. are present in the solution.   Where it can be safely assumed (*a*) that the change in the freezing point depression with concentration is a linear function of concentration within the small concentration change resulting from separation of solvent (solid) due to undercooling, (*b*) that no heat of dilution is involved, that is, if it can be assumed that the solution is acting as an ideal solution, and (*c*) that the solid separating consists only of the pure solvent component, correction of the observed freezing point depression, Δ', to yield the true value for the original solution, Δ, may be obtained by the relationship:

$$\Delta = \Delta' \left(1 - \frac{su}{\lambda a} \right)$$

where s = specific heat of the solution, u = the observed undercooling in °C., λ = latent heat of fusion of the pure solvent, and a = weight of solvent in a gram of solution. Since, for actual solutions, the above assumptions are only approximate under the most favorable circumstances, it is necessary that undercooling be avoided for most accurate results. Since undercooling is almost impossible to avoid, and frequently, when body or tissue fluids are involved, may be of large magnitude, it is obvious that this method should not be considered more than approximately exact. Unless the volume of solution employed in the determination is large, further corrections would have to be applied (in case of undercooling) for the specific heat of the mercury in the thermometer bulb, for the specific heat of the stirrer wire used, and probably for the energy added by the mechanical process of stirring while the solution warms up from the temperature of undercooling to the final temperature at which Δ' is read.

A second method for estimating freezing point depressions of dilute solutions, susceptible to considerably greater accuracy, is one introduced by Adams (12–14). In this method, samples of the pure solvent and of a solution are placed in highly lagged vessels (Dewar flasks) and brought to temperature equilibrium with the solid phase of the pure solvent present in each. The temperature difference is measured very accurately by use of a thermocouple (or thermopile) and at this stage a sample of the solution is removed for analysis. The accuracy of this method is dependent upon the exactness with which the temperature difference can be estimated and the accuracy of the analysis of the equilibrium sample of solution. With careful work the temperature difference may be obtained within an error of perhaps 0.0025 °C. For use in the determination of the freezing point depression of a tissue fluid such as blood serum, an analysis for some component (e.g., chloride) of the original sample of serum would have to be made. The serum would then have to be concentrated by removal of water and a series of freezing point determinations made such that the equilibrium concentrations (again followed by chloride determinations on the equilibrium samples) studied would span a range on either side of that of the original sample. A plot of the observed values of Δ versus chloride concentration, with interpolation to the chloride concentration of the original serum, should yield as

accurate a value of Δ for the original serum as is obtainable by any freezing point method. A disadvantage of the use of freezing point methods, generally, when biological fluids are being examined, is the fact that relatively large volumes of fluid must be employed.

Vapor Pressure Methods. While many methods have been devised for the determination of the molecular weights of solutes by measuring their effects upon the relative vapor pressures of the solvents, the accuracy claimed for these methods is in general not very high even when carried out, as is required for highest accuracy, at or near the boiling points of the solvents. Such temperatures can seldom be employed in biophysical measurements. The vapor pressure method that has attained the highest degree of precision and one that was specifically devised for use with biological tissues and fluids is the so-called thermoelectric method of Hill (15). The apparatus as modified by Baldes (16) consists of a sensitive thermocouple, the thermojunctions of which are made in the form of loops. Upon one loop is placed a drop of the sample, e.g., blood serum of unknown relative vapor pressure, and upon the other loop is placed a drop of a reference salt solution (of known relative vapor pressure), which is also used to moisten a filter paper lining the air-tight chamber in which the thermocouple is placed. The whole assembly is placed in a water bath of constant temperature. Condensation onto (or evaporation from) the drop of unknown will cause the thermojunction upon which it rests to become warmer (or cooler) than the other junction and the temperature difference is determined by the deflection caused in a sensitive galvanometer connected to the leads of the thermocouple. The instrument is calibrated by using a series of solutions of known relative vapor pressures, placing them successively on the junction later to be occupied by the unknown. The principle involved is, thus, essentially the same as that of the wet bulb thermometer in the determination of relative humidity. Roepke and Baldes (17) have studied this method and emphasize that, because it is a dynamic method (readings depend upon the rate of evaporation or condensation upon the junctions), any or all of the following circumstances may lead to error in determinations in which tissues or body fluids are used as unknowns in an instrument calibrated with simple electrolyte solutions: (a) surface films, (b) difference in coefficient of diffusion of water in sample from that in reference solution, (c) greater nonsolvent volume in sample than in reference solution, (d) difference in shape of drops, owing to surface tension differences,

(e) difference in heat of condensation, (f) thermochemical reactions in sample (particularly in intact cells), and (g) presence of volatile solutes. While the possible importance of these sources of error will vary with the nature of the sample being studied, these authors concluded that, if the "osmotic pressure of a sample of blood differed from that of the reference solution by 10 per cent, the osmotic pressure as determined with the thermocouple would be in error by less than 0.2 per cent."

This method, in the hands of a careful and meticulous worker, probably constitutes the best available means for determination of the osmotic pressure of solutions, containing low molecular constituents, when the determination must be made on the solution without change in concentration and at temperatures above the freezing point of the solvent. For example, the state of aggregation or dissociation of detergents in water solution is a function both of the concentration and of the temperature of the solution, there existing both molecular and micellar components of the detergent in the solution. Measurements made with the freezing point depression method yield information with regard to the state of the solute only at or near the freezing temperature of the solvent. Effects of change in temperature of the solution upon the dissociation–association properties of the solute cannot be studied by this method. Direct osmotic pressure measurements are very difficult to make on such systems because of the difficulty in obtaining semipermeable membranes that will be impermeable to the low molecular components. This vapor pressure method can be employed on such systems with considerable success (18). It is quite probable that similar variations with concentration and temperature may obtain in biological systems. This method, when the measurements are made at body temperature, eliminates the possibility of errors arising from these possible variations.

2. Direct Methods

Plasmolysis Method. Where isolated cells or small tissue specimens are being investigated, the simplest and most direct method for determining their osmotic properties is by the plasmolysis method commonly attributed to de Vries. By placing specimens in a series of solutions of varying concentrations of some solute or mixture of solutes, the osmotic pressures of these solutions being known, and observing, with a microscope, which solutions cause the

least change in volume of the specimen, the osmotic pressure of the
solution contained within the cells can be obtained.   Solutions of
osmotic pressure greater than that of the specimen contents (hy-
pertonic solutions) will cause shrinkage (plasmolysis) of the cells
due to passage of water out of the cell interior while solutions of os-
motic pressure less than that of the specimen (hypotonic solution) will
allow swelling of the cells.   That solution with the same osmotic
pressure as the cell contents (isotonic solution) will cause no change
in its volume.   It is essential to bear in mind that the validity of the
results of such determinations will depend upon the efficiency with
which the cell membrane acts in a semipermeable manner both to the
solutes contained in the cell and to the solute components of the test
solution.   This is usually a point of considerable uncertainty.   If the
membrane were permeable to solvent but completely impermeable
to all solutes both within and outside the cell, the relative activity of
the solvent on the two sides of the membrane would be the sole de-
termining factor for the passage of solvent.   On the other hand if
the membrane were completely permeable to all molecular species
both solutes and solvent, and if the diffusion constants of the solute
components in both solutions were of about the same magnitude,
again little or no net transfer of solvent would occur when the two
solutions were isotonic.   If, however, the membrane possesses a dif-
ferential permeability to various solute components or if the diffusion
constants of various solute components were very different, some
transfer of solvent would occur even though the two solutions bathing
the opposite sides of the membrane were initially isotonic.   It can
generally be assumed that water will move across the membrane with
greater ease than will any of the solutes, so that, if the observations
are based on the initial trend, more reliance might theoretically be
placed upon them than when the observations are made at equilib-
rium.   A disturbing circumstance exists, when initial changes in
volume are being employed in such determinations, in that membrane
potentials (electrical) that may occur across the membrane can,
through electroosmotic effects, constitute the most important factor
determining the initial passage of water across the membrane (21).
Anomalous osmosis resulting from the performance of work by ex-
penditure of electrical energy in such cases would destroy the validity
of these results when interpreted as due to osmotic forces alone.
Many cell membranes, however, appear to act as truly semiperme-
able when the solutes of the test solution are not greatly divergent

in nature from the solutions bathing the cell in its natural habitat, and equilibrium measurements may be reliable in such cases. Some cells are so perfect as osmometers, their membranes being readily permeable to water but impermeable to other small molecular species, that they have been employed successfully in the determination of the molecular weights of solutes. De Vries was able to determine the molecular weight of raffinose by plasmolysis experiments with the cells of *Tradescantia discolor*. Overton compared the concentrations of solutions of sugars and amino acids that would just cause plasmolysis of the cells of *Spirogyra* and found them to be closely proportional to the molecular weights of these solutes. The subject has been reviewed by Lucké and McCutcheon (*19*) and Lucké (*20*).

Osmometers. The direct determination of the osmotic pressure of a solution for the purpose of estimating molecular weights of solutes, must rely, for dependability, upon the use of a membrane which is impermeable to the solute the molecular weight of which is desired, but which is readily permeable to the solvent and to any other solute constituents that may be present in the solution. Membranes should be of uniform properties as to thickness and porosity and subject to easy and uniform reproducibility. The extensive investigations of Manegold (*22*) have furnished much information on the production of such membranes. Collodion membranes can be readily prepared in the forms required for use in the various osmometers described, and such membranes have been found generally satisfactory for measurements of the osmotic properties of high molecular substances in aqueous solutions. Cellulose membranes regenerated from cellulose nitrate or viscose are obtainable commercially (as unwrinkled material that has never been dried or treated for waterproofing) or may be regenerated from cellulose nitrate by a method such as that described by Montonna and Jilk (*23*). Such membranes have been used successfully for solutions in organic solvents. A primary limitation to the use of any such membranes is that the molecular size of the solute under study must be great enough so that the membrane will not allow diffusion of the solute through its pores. If the porosity of the membrane is not known, analysis for the solute in the external phase, at the end of a determination, should always be made to be certain that the membrane used has acted dependably in this regard. Carter and Record (*40*) describe a method by which commercial nonwaterproofed viscose can be prepared as membranes of various degrees of porosity for use in aqueous

or organic solvents. From the standpoint of the time required to make the measurements, it is advantageous to use a membrane with the highest obtainable degree of porosity for the solvent, so long as it is completely impermeable to the solute under consideration.

The great profusion of osmometers described in the literature can be classified broadly into two types. Those that measure, at equilibrium, the pressure difference developed across the membrane in a liquid column or head resulting from an influx of solvent into the solution are called *static elevation* osmometers. Those that measure the instantaneous diffusion pressure of the solvent through the application of an external pressure to the solution side (or a negative pressure to the solvent side) of the membrane which is just sufficient to prevent flow of solvent into the solution are called *dynamic equilibrium* osmometers. Further variations in the construction and mode of operation of the osmometers employed by various investigators are due primarily to (*a*) the shape of membrane employed, (*b*) the magnitude of the pressures that they are designed to measure and the devices used to measure the pressures, (*c*) the volumes of solution and solvent required to fill them, and (*d*) the manner in which surface tension effects are eliminated or corrected for. Arrangements for stirring the solutions to hasten equilibrium and attachments designed to minimize evaporation are often included in the design of the osmometers, particularly in the static elevation type.

The static elevation osmometer has the advantage of simplicity of construction and operation but the disadvantage that the length of time required for making a measurement is relatively great. Since the pressure difference is developed by a movement of solvent into the solution, the concentration of the latter will be changed during the course of equilibration. This is usually minimized by starting the determination with a pressure on the solution approximately equal to the equilibrium pressure expected (determined usually by a preliminary run) and by using a capillary tube in which the column of liquid giving rise to the pressure is allowed to develop. Measurement or calculation of the surface tension pull exerted on the liquid in the capillary must be taken into account in obtaining the effective pressure at the membrane. Allowance for density differences between the solution and solvent must likewise be made. The use of a flat membrane leads to a somewhat greater degree of accuracy in estimating the active pressure than is the case with the saclike membrane, especially where very small osmotic pressures are being measured.

There exists a wide variety of choice among the static elevation types of osmometers described in the literature (*32–47*).

The dynamic equilibrium osmometer, first used by Berkeley and Hartley in their studies on sucrose solutions in which they obtained results in accurate agreement with those obtained by the static elevation method of Morse and Frazer with the same solute, has the advantage of rapid determination of the osmotic pressure, but this type of instrument is somewhat more difficult to fill and operate than are most of the static elevation apparatus. It has the limitation also that, unless the activities of all *diffusible* components other than the solvent are equal across the membrane when the osmotic pressure is measured, the values of osmotic pressure so obtained are of no value in the calculation of the molecular weight of the nondiffusible component. In those cases, however, in which it is known that no diffusible component other than the solvent is present in the system, or where it is known that the activities (concentrations) of such diffusible components are exactly equal in the two solutions on the opposite sides of the membrane, this method allows for a much quicker estimation of the osmotic pressure due to the nondiffusible component than does the static method. Many of the static elevation type osmometers utilize an applied constant external pressure approximately equal to the osmotic pressure to be expected at equilibrium, thus minimizing the degree to which solvent transfer across the membrane must occur during the attainment of the final equilibrium. Most dynamic equilibrium type osmometers can be readily adapted in their use to the static elevation procedure. Descriptions of a number of osmometers employed primarily in the dynamic equilibrium procedure can be found in references given under this heading (*24–31*).

Since the type and specific construction of the osmometer that will prove best adapted to a particular problem will necessarily vary with the properties (molecular size, whether electrolyte or nonelectrolyte, solubility, amount available, etc.) of the solute, it is recommended that the investigator, in light of his particular problem, consider in detail the characteristics of a number of the instruments to which reference has been made. When the objective of the experiment is to obtain the molecular weight of a high molecular material from osmotic pressure measurements, it is necessary to bear in mind that some form of extrapolation to very low or zero concentration is almost always necessary. The osmometer selected must be capable, therefore, of yielding accurate values of osmotic pressure through a

range of pressures and at sufficiently low concentrations of solute to make such an extrapolation dependable. Osmotic pressures as low as, or even lower than, 10 mm. of water will often be necessary. At such low pressures, surface tension differences between solvent and solution, and their effects on the capillary rise of the liquids in the capillary tubes that may be employed for creating the hydrostatic head at equilibrium, must be taken accurately into account as must also density differences between the solutions. The osmometer should be so constructed that it can be held in a constant temperature bath accurate to 0.02 to 0.01 °C. in order to eliminate as far as possible fluctuations due to density changes and surface tension changes as well as osmotic pressure changes with temperature.

References

General References

1 . Lewis, G. N., and M. Randall. *Thermodynamics*. McGraw-Hill, New York, 1923.
2 . Mark, H., *Physical Chemistry of High Polymeric Systems* (High Polymers. Vol. II). Interscience, New York, 1940, p 228.
3 . Wagner, R. H., "Determination of Osmotic Pressure," in *Physical Methods of Organic Chemistry*, 2nd ed., A. Weissberger, ed. Interscience, New York, 1949, Chap. XI.
4 . Ferry, J. D., "Ultrafilter Membranes and Ultrafiltration," *Chem. Revs.*, **18**, 373 (1936).
5 . Lifson, N., and M. B. Visscher, "Osmosis in Living Systems," *Medical Physics*, O. Glasser, ed. Year Book Publishers, Chicago, 1944, p. 869.

Theory

6 . Ostwald, W., *Kolloid-Z.*, **49**, 60 (1929).
7 . Flory, P. J., *J. Chem. Phys.*, **10**, 51 (1942).
8 . Huggins, M. L., *J. Am. Chem. Soc.*, **64**, 1712 (1942); *J. Phys. Chem.*, **46**, 151 (1942); *Ind. Eng. Chem.*, **35**, 216 (1943).
9 . Scatchard, G., *J. Am. Chem. Soc.*, **66**, 2315 (1946).
10 . Adair, G. S., *Proc. Roy. Soc. London*, **A109**, 292 (1925); **A120**, 573 (1928); **A126**, 16 (1929); *J. Am. Chem. Soc.*, **51**, 696 (1929).
11 . Donnan, F. G., *Chem. Revs.*, **1**, 73 (1924).

Methods and Techniques

FREEZING POINT

12 . Adams, L. H., *J. Am. Chem. Soc.*, **37**, 481 (1915).

13. Randell, M., and A. P. Vanselow, *J. Am. Chem. Soc.*, **46**, 2418 (1924).

14. Hovorka, F., and W. H. Rodebush, *J. Am. Chem. Soc.*, **47**, 1614 (1925).

VAPOR PRESSURE

15. Hill, A. V., *Proc. Roy. Soc. London*, **A127**, 9 (1930).

16. Baldes, E. J., *J. Sci. Instruments*, **11**, 223 (1934); *Biodynamica*, No. 46 (1939);   E. J. Baldes and A. F. Johnson, *Biodynamica*, No. 47 (1939). See also R. W. Culbert, *J. Biol. Chem.*, **109**, 547 (1935).

17. Roepke, R. R., and E. J. Baldes, *J. Biol. Chem.*, **126**, 349 (1938).

18. Fineman, M. N., and J. W. McBain, *J. Phys. & Colloid Chem.*, **52**, 881 (1948).

PLASMOLYSIS

19. Lucké, B., and M. McCutcheon, *Physiol. Revs.*, **12**, 68 (1932).

20. Lucké, B., *Cold Spring Harbor Symposia Quant. Biol.*, **8**, 123 (1940).

MEMBRANES

21. Loeb, J., *J. Gen. Physiol.*, **1**, 717 (1919);   **2**, 173, 255, 387, 563, 577, 659 (1920).   See also K. Sollner and H. P. Gregor, *J. Phys. & Colloid Chem.*, **50**, 470 (1946);   **51**, 300 (1947).

22. Manegold, E., *Kolloid-Z.*, **61**, 140 (1932).

23. Montonna, R. E., and L. T. Jilk, *J. Phys. Chem.*, **45**, 1376 (1941).

OSMOMETERS, DYNAMIC EQUILIBRIUM TYPE

24. Berkeley, Earl of, and E. G. J. Hartley, *Trans. Faraday Soc.*, **A206**, 486 (1906).

25. Sørensen, S. P. L., *Z. physiol. Chem.*, **106**, 1 (1919).

26. Herzog, H. O., and H. M. Spurlin, *Z. physik. Chem.*, *Bodenstein Festband*, 239 (1931).

27. van Campen, P., *Rec. trav. chim.*, **50**, 915 (1931).

28. Obogi, R., and E. Broda, *Kolloid-Z.*, **69**, 172 (1934).

29. Boissonnas, C. G., and K. H. Meyer, *Helv. Chim. Acta*, **20**, 783 (1937).

30. Meyer, K. H., E. Wolff, and C. G Boissonnas, *Helv. Chim. Acta*, **23**, 430 (1940).

31. Fuoss, R. M., and D. J. Mead, *J. Phys. Chem.*, **47**, 59 (1943).

OSMOMETERS, STATIC ELEVATION TYPE

32. Morse, H. M., and J. C. W. Frazer, *Am. Chem. J.*, **34**, 28 (1905);   **38**, 122 (1907).

33. Frazer, J. C. W., *J. Am. Chem. Soc.*, **38**, 1907 (1916);   **43**, 2497 (1921); **45**, 1710 (1923).

34. Adair, G. S., *Proc. Roy. Soc., London*, **A108**, 627 (1925).

35. Burk, W. F., and D. M. Greenberg, *J. Biol. Chem.*, **87**, 197 (1930).

36. Dobry, A., *J. chim. phys.*, **32**, 46 (1935).

37. Oakley, H. B., *Trans. Faraday Soc.*, **31,** 136 (1935).

38. Keys, A., and H. Taylor, *J. Biol. Chem.*, **109,** 47 (1935).

39. Bourdillon, J., *J. Biol. Chem.*, **127,** 617 (1939).

40. Carter, S. R., and B. R. Record, *J. Chem. Soc.*, **1939,** 660.

41. Schulz, G. V., *Z. physik. Chem.*, **A176,** 317 (1936); **A180,** 1 (1937); **B52,** 1 (1942).

42. Bull, H. B., *J. Biol. Chem.*, **137,** 143 (1941).

43. Gee, G., *Trans. Faraday Soc.*, **36,** 1162, 1171 (1940); **38,** 147 (1942).

44. Flory, P. J., *J. Am. Chem. Soc.*, **65,** 372 (1943).

45. Gee, G., in *Advances in Colloid Science*, Vol. II. Interscience, New York, 1946, p. 151.

46. Ewart, H. R., *ibid.*, p. 220.

47. Scatchard, G., S. C. Batchelder, and A. Brown, *J. Am. Chem. Soc.*, **66,** 2320 (1946).

CENTRIFUGATION

E. G. Pickels, *Specialized Instruments Corporation*

A. FUNDAMENTAL PRINCIPLES

1. Sedimentation in Uniform Field of Force

Under the influence of gravity, particles suspended in a liquid of lower density tend to settle to the bottom of the containing vessel. If there were no other disturbing influences, each particle would sediment at a rate governed primarily by its size, shape, and density. If all the particles in a very dilute suspension were identical, they would sediment at the same rate, and if in a vertical column of fluid the starting concentration were uniform, those particles originally at the surface would form a sharply defined "rear line of march" or moving boundary, which would demarcate the supernatant fluid and the sedimenting phase.   Until the boundary has reached the bottom of the vessel, the number of particles piling up there would increase at a regular rate, while the concentration in the region between the boundary and the bottom of the vessel would remain uniform and constant.   If a preparation under such conditions contained two groups of particles of two different sedimentation rates, there would be formed two different boundaries, the increase in the concentration of the suspended phase at each boundary representing the concentration of the respective component.   In general, multiple boundaries would represent discrete particulate components, and in typical cases would probably be associated primarily with differences in particle size.   On the other hand, a group of particles having a continuous gradation in size, shape, or density within a limited range would sediment at varying rates and thus exhibit a spreading or blurring of the boundary about the mean position, the amount of spread increasing in a regular manner with the displacement of the boundary from the meniscus.

In actual practice, a boundary representing a monodisperse system is not infinitely sharp because of the thermal agitation or Brownian movement of the particles. The boundary becomes progressively less well defined with time because of this superimposed diffusion across the boundary.   Also, the particles approaching the bottom of the vessel do not all proceed in a regular fashion to be incorporated into a closely packed sediment.   There is no sharp demarcation between the sediment and the sedimenting phase, but rather a transition zone in which the concentration is graded from that of the sedimenting to that of the sedimented phase.   However, when the rate of sedimentation is rapid enough in comparison to the diffusion proc-

ess, there exists for some time, between the diffuse boundary and the transition zone, a region ("plateau" region) in which the concentration remains nearly constant and uniform throughout. This circumstance permits one to obtain a measurement of the true sedimentation rate in spite of the diffusion, for it is known from the nature of the diffusion process that the position which the boundary would have had in the absence of diffusion is simply the level at which the concentration is one-half that in the plateau region (*1*, p. 6).

Practical measurements that can be completed in a matter of hours with gravitational force alone are in general limited to particles such as red blood cells, etc., in the size range of a few microns or more. The boundary spreading due to diffusion is for such particles negligible in comparison to the boundary movement and hence appreciable inhomogeneity in sedimentation rate is immediately apparent from a lack of sharpness in the boundary. Also, if there are several components of discrete particle sizes, their respective boundaries are relatively sharp and easily differentiated.

2. Sedimentation in Centrifugal Field of Force

The sedimentation behavior of materials in the size range considerably below a micron, such as virus particles and protein molecules, cannot be studied with gravitational force alone. As a general rule, the smaller the particle size, the slower the sedimentation and the more pronounced the diffusion process. When the diffusion rate is sufficiently high in comparison to the sedimentation rate, a boundary does not become well enough differentiated for analysis. In containing vessels of practical size, no distinct supernatant zone relatively free of particles becomes established before the diffuse boundary begins to grade into the transition zone with the consequent loss of any well differentiated plateau region. Furthermore, even if diffusion were not a complicating factor, the time required for a determination of sedimentation rate would be too long for practical purposes in most cases.

The use of high centrifugal forces instead of gravity permits application of the method to relatively small particles, although the range is still limited by the same basic considerations. For example, although modern ultracentrifuges are capable of developing centrifugal forces in excess of 250,000 times gravity, application of the boundary method with any degree of precision is still limited in the case of monodisperse preparations to materials having molecular

weights above several thousand. In the case of paucidisperse prepa-
rations that would show multiple boundaries in the absence of dif-
fusion, molecular weights generally must be well above 10,000 to ac-
complish noticeable differentiation. Also, the smaller the particle
size, the greater the uncertainty of differentiating the amount of
boundary spreading due to diffusion and the amount due to inhomo-
geneity of sedimentation rate, and the greater the necessity of acquir-
ing separate diffusion data for making the comparison. Generally
speaking, diffusion can hardly be considered a negligible factor in the
analytical centrifuge until a size range at least as high as that of the
larger viruses (above 50 mμ roughly) is reached.
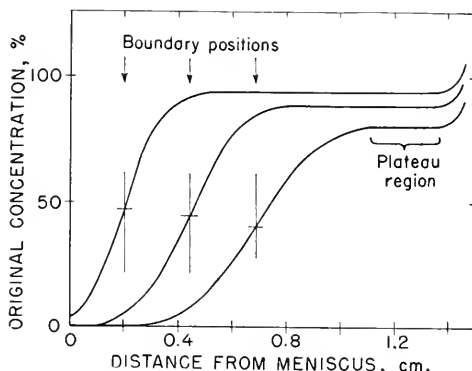
Fig. 1. Three curves illustrating the concentration
distribution of slowly sedimenting material at eighty-
minute intervals after boundary has cleared meniscus
sufficiently to permit measurements of position.

As illustrated in Figure 1, the general character of the sedimenta-
tion process in the centrifuge and the method of analysis are similar
to those described for a gravitational field, except for certain condi-
tions imposed by the nonuniformity of the centrifugal field. Since
the field is radial, the containing vessel or cell in an analytical centri-
fuge is made sector-shaped (20), having two flat walls that, if ex-
tended, would intersect along the axis of rotation, while the other
two side walls are perpendicular to the axis. Thus, particles origi-
nally close to any of these walls continue during sedimentation to pur-
sue an average course parallel to the wall, so that there is no sedimen-
tation against or away from the walls to introduce convective dis-
turbances. Also, since in their sedimentation the particles on the

average follow diverging radial paths and since their rate of migration increases with the distance from the axis of rotation, the concentration in the plateau region steadily decreases. However, the true boundary position at any time is still very nearly the level at which the concentration is half that of the plateau region at that time. Furthermore, there is a definite mathematical relationship (see Sec. C) between the concentration of the plateau region and the displacement of the boundary, so that the initial concentration of any differentiated component can still be determined by applying a correction factor.

The curves in Figure 1 correspond to behavior of a typical "globular" protein of 18,000 molecular weight in average centrifugal field of 300,000 times gravity. To first approximation, a similarly shaped larger protein sedimented to corresponding positions by the same force would show reduced boundary spreading inversely proportional to the two-thirds power of the molecular weight. The time required for sedimentation would be lowered in the same ratio.

3. Significance of Ultracentrifugal Method

Although sedimentation rate is not of itself an indication of particle size, shape, or density, it is for a given preparation under appropriate conditions a characteristic constant that may be combined with other experimental data to investigate these properties. Although average sedimentation rates may be determined from the proportional amount of material sedimented to the bottom of or across a certain level of the containing vessel within a known time, the moving-boundary method is unique in that through a single experiment one can, with proper recording equipment (usually photographic), study the distribution of sedimentation rates and thus the number, centrifugal homogeneity, and respective concentrations of the resolvable components. For the perfection of this method in the form of the analytical ultracentrifuge, and for numerous demonstrations of its practicability in the study of macromolecules, science is indebted to the pioneering research of Svedberg and associates (1,9,13) begun about 1923.

As already indicated, no measurable sedimentation boundary is established with diffusible materials in centrifuge cells of ordinary size when the centrifugal force applied is below a certain value usually dependent primarily on the particle size. However, if steady centrifugation is continued for a sufficient length of time (at least one day for most proteins), an equilibrium is established between sedimenta-

tion and the backward diffusion.   In this steady state condition the
concentration grades from a value below the original at the meniscus
to values higher than the original in the outer zone of the cell.   It has
been shown by Svedberg and associates (*1*) that this concentration
distribution is a function of particle weight and can be employed
directly (see Sect. C) for its determination in the case of monodisperse
preparations.   It has been demonstrated (*27*) that even the simpler
molecules with molecular weights below 100 can be studied by this
equilibrium method.   However, its principal application has been as
an independent method for investigating proteins and other macro-
molecules, for which purpose it offers certain advantages by reason of
the simultaneous balancing of the sedimentation and diffusion proc-
esses under identical experimental conditions.

B. ULTRACENTRIFUGES OF OPTICAL TYPE

1. Ultracentrifuges of Vacuum Type

Although the pioneering oil-driven ultracentrifuges of Svedberg
and collaborators (*1,13,59*) are still in active and fruitful use at Upsala
and several other research institutions, most present day machines
are of the somewhat simpler and more versatile vacuum type.   A
commercial ultracentrifuge (*89*) of this kind suitable for both sedi-
mentation velocity and sedimentation equilibrium measurements,
as well as for preparative work, is available.   The complete instru-
ment is slightly more than 6 ft. long, weighs approximately 2400 lb.,
and requires no special foundation by virtue of its own system of vibra-
tion isolation.   Size and other pertinent features of the analytical
rotor and its transparent cell have been patterned after those found
to represent the optimum by Svedberg (*1*).   The rotor (Figs. 2 and
3), with a maximum diameter of 7.25 in., is made from a solid piece of
forged Duralumin and is provided with two 1-in. holes located 65
mm. from the axis of rotation and accommodating the solution cell
and its counterbalance.   Metal is cut away to give the oval shape,
relieving the stresses in the plane of the cell holes and permittng
higher speeds (*6*).   The rotor may be operated routinely at speeds up
to at least 60,000 r.p.m., with average centrifugal forces in the cell
up to 260,000 times gravity.

The cell (*7*) (see Fig. 2) is composed primarily of a Duralumin center-
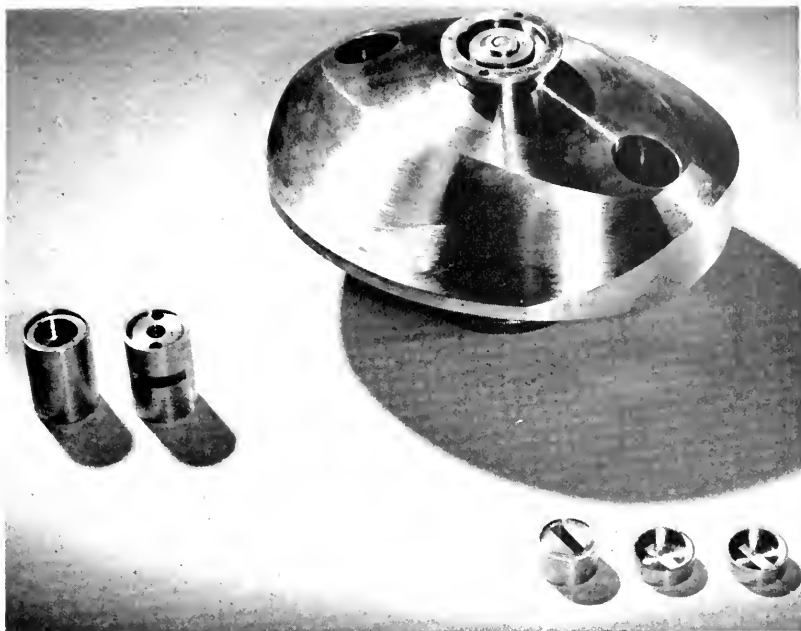piece, two quartz discs, two light-limiting apertures of Duralumin (which

Fig. 2. Analytical rotor for a commercially produced, electrically driven ultracentrifuge (89) of the optical type. Also shown are assembled cell, counterbalance, separate centerpiece for fluid column, and sector cups that hold two quartz windows. (Courtesy *Specialized Instruments Corporation*.)

Fig. 3. Analytical rotor of Figure 2 shown attached to flexible drive shaft in operating position. Vacuum chamber and inner refrigerated cylinder shown in lowered position. (Courtesy *Specialized Instruments Corporation*.)

serve as window receptacles also), a threaded Duralumin barrel or casing, and a screw ring of the same material. The centerpiece, which is generally anodized for chemical resistance, is provided with a sector-shaped slot, 15 mm. long (in direction of centrifugal force) and a few millimeters wide, that accommodates the fluid. Surrounding the central slotted section are concentric grooves to promote a fluid-tight seal against thin washers of plastic placed between the centerpiece and the quartz windows. The assembled cell is filled with the aid of a hypodermic syringe at the narrow end of the slot through a small hole, which is then sealed with a thin rubber gasket and a small screw plug fitting a threaded hole in the barrel. Centerpieces with fluid column thicknesses (axially) up to more than 12 mm. may be used. Light passing through a hole in the counterbalance at a radial distance greater than that corresponding to the bottom of the cell slot furnishes a reference for the light intensity or the distance from the axis of rotation.

The rotor spins about a vertical axis within an evacuated chamber, seen in the lowered, opened position in Figure 3. The double-walled steel cylinder and thick steel end plates, along with a heavy locking device and the steel walls of a surrounding barricade, furnish protection to the operator and the equipment in the event of rotor failure. A third, inner cylinder, which immediately surrounds the rotor, can be refrigerated for work at reduced temperatures. Pressures well below 1 μ of mercury are obtained in the vacuum chamber by means of an oil diffusion pump backed up by a mechanical vacuum pump. The rotor is attached by a coupling device to a spring steel shaft 0.1 in. in diameter, which passes through a vacuum-tight oil gland to the driving mechanism above the vacuum chamber. Suspension of the rotor by means of a flexible shaft makes the rotor self-balancing and permits an off-balance of several grams at the cell without excessive vibration or other ill effects. The arrangement is an advantage in the event of cell leakage and also avoids the necessity of extreme accuracy in balancing the rotor. There is an appropriate damping device to prevent swinging or precession of the rotor and a collecting system for the few milliliters of oil that escape through the sealing gland during a run. The temperature of the rotor can be measured with a thermocouple attached to the rotor before and after a run, and also during a run by a radiation couple located near the coupling device. At a speed of 60,000 r.p.m., the rise in temperature of the rotor is of the order of 1 °C. per hour if the surroundings are at a temperature near that of the rotor. By lowering the temperature of the refrigerated cylinder about 10 or 15 °C. below that of the rotor, one can keep the rotor temperature nearly constant indefinitely.

The driving mechanism consists of a series-wound, brush type electric motor (115 v.), a high speed gear train, and a system of specially designed mechanical bearings. The motor, cooled by circulation of both air and water, develops 1.5 horsepower and will get the centrifuge to full speed within seven minutes. Deceleration of the rotor to rest is accomplished in a com-

parable time by reversing and regulating the current through the armature. Accurate speed control at any of thirty different values is obtained by matching a greatly reduced speed from the drive against a selected speed controlled by a constant speed, synchronous motor.   Any difference motivates a differential gear, which in turn actuates the electrical controls that regulate the power supply to the drive motor.   The speed is also indicated by an electrical tachometer.

To permit viewing and photographic registration of the sedimentation according to methods discussed later, light from a mercury arc is directed upward through the revolving cell and through windows (which serve as collimating lenses also) in the vacuum chamber against a 45° mirror near the top of the instrument and thence horizontally through appropriate lenses to the photographic plate at the right end of the machine.   By means of a half-reflecting mirror, some light is diverted to a viewing screen so that the sedimentation pattern may be viewed at any time, even while photographs are being taken.   The photography is automatic, there being appropriate adjustments for preselecting both exposure time and the interval between photographs.   Exposure time is usually of the order of fifteen seconds.

A smaller (half size), relatively simple air-driven vacuum type ultracentrifuge patterned after that described by Beams and Pickels (17) has been commercially produced (91); its resolving power, however, is rather low (1, p. 47).

Since the description of the first air-driven centrifuge of the vacuum type was published in 1935 (19a), several improved designs for driving mechanisms have been described in the literature.   Descriptions of or reference to most of these can be found in the publications of Beams (8) and Pickels (2,4).   Of particular interest are the electrically driven (high frequency induction type), magnetically supported drive of Skarstrom and Beams (10) and the air-driven, air-supported "turret" type drive of Pickels (11), which has been in successful use at a number of research institutions for several years.

2. Svedberg Ultracentrifuges

The original Svedberg ultracentrifuge was a small "optical centrifuge" developed by Svedberg and Nichols (9) at the University of Wisconsin in 1923.   Later Svedberg and Rinde (20) used the machine to study size distribution among gold sols and proposed the name "ultracentrifuge" denoting an instrument by means of which sedimen-

tation in a centrifugal field could be measured quantitatively.   However, it has become common practice to associate the term "ultracentrifuge" with any type of high speed centrifuge rather than with the original connotation of quantitative measurement.

In 1927 Svedberg and Lysholm (12) substituted an oil drive for the electrical drive.   In the present model (1,13,90) used for sedimentation velocity measurements a rotor of special steel is driven by oil under pressure about a horizontal axis and in an atmosphere of hydrogen gas at a pressure of about 20 mm.   The hydrogen gas is necessary for conducting away from the rotor the comparatively large amount of heat developed through friction in the mechanical bearings. Fortunately, from the standpoint of avoiding temperature gradients (which would produce convection currents) across the cell in an axial direction, the rotor is symmetrical about a plane perpendicular to the axis, having a similar bearing and turbine assembly on either side. The temperature of the rotor is measured by a radiation thermocouple placed about 0.25 mm. from the rotor near the cell hole position.   Speed is determined by a stroboscope or a frequency meter. With certain special experiments small rotors with a cell height of 8 mm. and an average radial distance of 3.25 cm. have been operated at speeds up to 140,000 r.p.m.   However, it has been found that for general purposes the most practical rotor is one capable of operating at speeds up to about 65,000 r.p.m. with a fluid column approximately 17 mm. long at about 65 mm. from the axis of the rotation.   Fluid column thicknesses up to about 12 mm. are used according to the concentration and nature of the solute.   Cell distortion, leakage, and cell failure become serious problems as centrifugal forces considerably above 300,000 times gravity are employed.   Since the rotor shaft and the bearings are comparatively rigid in the oil-driven ultracentrifuge, dynamic balance is very critical, although a damping device (1) used in the later models has reduced vibration somewhat.

For use as an equilibrium centrifuge Svedberg and Sjögren (14) in 1929 improved the early low speed electrical centrifuge by substituting a direct motor drive for the earlier gear drive.   The drive rotates about a vertical axis and is of a self-balancing type used for spinning viscose thread in the artificial silk industry.   It is fed with three phase a.c. current of variable frequency from a special generator and has a short-circuited squirrel cage rotor moving in ball bearings. The 15 cm. rotor spins in an atmosphere of hydrogen and has an accurate temperature control.   The transparent cell and counterbal-

ance are of the same general character as those used with the oil-driven ultracentrifuges.   Speeds up to about 20,000 r.p.m. can be employed.   Machines of this type are now commercially available (90).

3. Optical Ultracentrifuges Exposed to Air Friction

Small optical ultracentrifuges, usually less than 5 cm. in diameter, have been constructed without mechanical bearings according to principles originally described by Henriot and Huguenard (15).   The rotor is cone-shaped at the base and rides on a whirling layer of air issuing under pressure from properly directed jets in a cone-shaped stator.   Beams, Pickels, and Weed (16) have pointed out the difficulties of obtaining normal sedimentation in such centrifuges because of the disturbing action of convection currents set up by temperature gradients through the rotor.   They were able to obtain sedimentation photographs of hemoglobin, but found that for this the thickness of the fluid column had to be kept below 1 mm. if reasonably dilute preparations were used.   McBain and Lewis (19b) and Beams (8) have made various improvements for minimizing the convection currents, and the former have described novel methods for photographically recording boundary positions.   These "spinning tops" are relatively inexpensive and could be used in many cases where more elaborate equipment is not justified.   However, from the standpoint of completely convection-free sedimentation at low concentrations, they do not approach the ideal of a thermally isolated system (as does the vacuum centrifuge) and because of practical limitations regarding the length, axial thickness, and radial displacement of the cell, their resolving power, precision of measurement, and range of application with respect to solute concentration do not compare very favorably with the ultracentrifuges already described (1, p. 47).

A small air-driven ultracentrifuge (92) of plastic material spinning about a horizontal axis on mechanical bearings has been described by Stern (18).   While probably sufficient for some studies, it is not suited to work with most proteins or with very dilute solutions because of its limited speed range and its susceptibility to convective disturbances.

4. Optical Methods

In the first analytical ultracentrifuges, measurements were based on the light absorption method of Svedberg and Rinde (20).   Filtered

light, usually ultraviolet, and of such wavelength as to be differen-
tially absorbed by solute and solvent, is directed through the revolving
cell.   Since the sedimentation occurs radially, the meniscus of the
fluid in the cell would appear to the eye as a fine line curved about the
axis of rotation.   The sedimenting boundary is also concentric about
the axis of rotation.   A record of the sedimentation is obtained by pho-
tographing at intervals only a narrow cross-sectional strip of the total
annular band (1).   A camera lens of long focal length (100 cm.) is
used to avoid errors of parallax and to give good depth of focus.   Typi-
cal photographic records are shown in Figures 4 and 5.
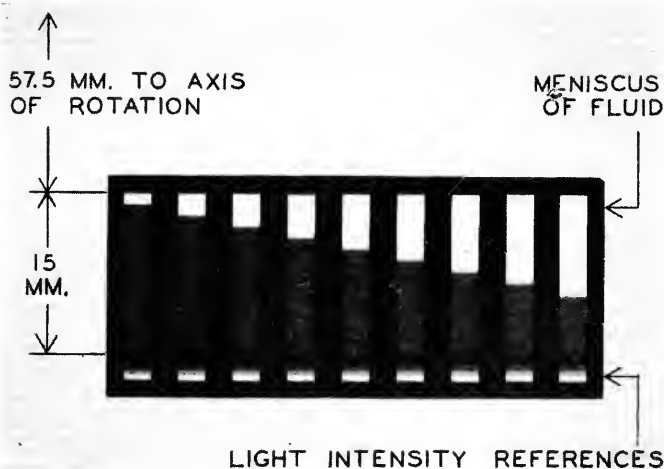
Fig. 4.   Sedimentation of a monodisperse, high molecular protein (hemo-
cyanin, approximate molecular weight 3.5×10^6) as recorded by the ab-
sorption method.   Photographs taken fifteen minutes apart with ultra-
violet light;   speed, 18,000 r.p.m.

The original absorption method has been largely superseded by
methods (4,21–24) based on the detection or measurements of devia-
tions suffered by light rays passing through regions of refractive index
gradient, such as exists at sedimenting boundaries.   The two principal
methods in current use are the Lamm scale method (24,26) and the
cylindrical lens method (4,23).   In the first of these a real or opti-
cally projected scale is placed between the revolving cell and a dif-
fuse source of illumination.   The equally spaced (not more than a few
tenths of a millimeter) lines of the scale are oriented tangentially with

respect to the rotor. The scale is photographed through the solution with a camera of long focal length. In the photographic image of the scale the lines are found to be still equally spaced in the regions corresponding to supernatant fluid and the plateau region. However, in the portion corresponding to a sedimenting boundary the lines are found to be displaced with respect to some reference line photographed through the opening in the counterbalance. The amount of displacement is proportional to the refractive index gradient and through plotting line displacement as a function of the distance from the axis of rotation one can obtain a curve that gives the refractive

REFRACTIVE   ABSORPTION
INDEX METHOD   METHOD
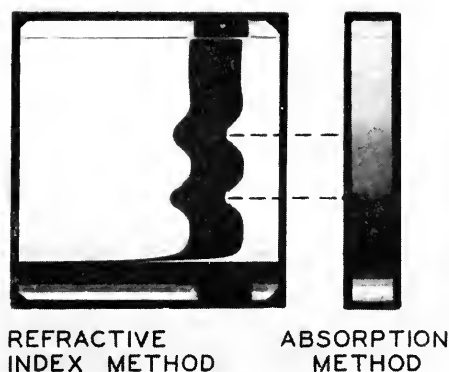
Fig. 5. Sedimentation photographs of dissociated hemocyanin showing double boundary and illustrating connection between absorption method and cylindrical lens method, which depends on refraction of light by concentration gradients. Speed, 16,000 r.p.m.; time at full speed, 2.25 hours; cell thickness, 1 cm.; protein concentration, 0.4%.

index gradient and hence the concentration gradient at all radial positions in the cell. The correction and interpretation of such curves are the same as those for the cylindrical lens method, which is discussed in detail below. The scale method when properly employed is probably subject to the fewest optical errors of any refractive index method. However, it does involve a great deal of tedious measurement and computation, which can be avoided through use of the cylindrical lens method.

In contrast to the scale method, the cylindrical lens method gives directly by photography a curve of refractive index gradient as a function of radial position within the cell. A schematic representa-

tion of the system used on the vacuum ultracentrifuge already de-
scribed and of the optical principles involved is given in Figure 6.
Consider the first arrangement shown, with the thin object diaphragm
placed between the collimating and condensing lenses.   All elements
are centered on an optical axis and disposed along it to fulfill the fol-
lowing conditions: light rays from a common point in the slit source
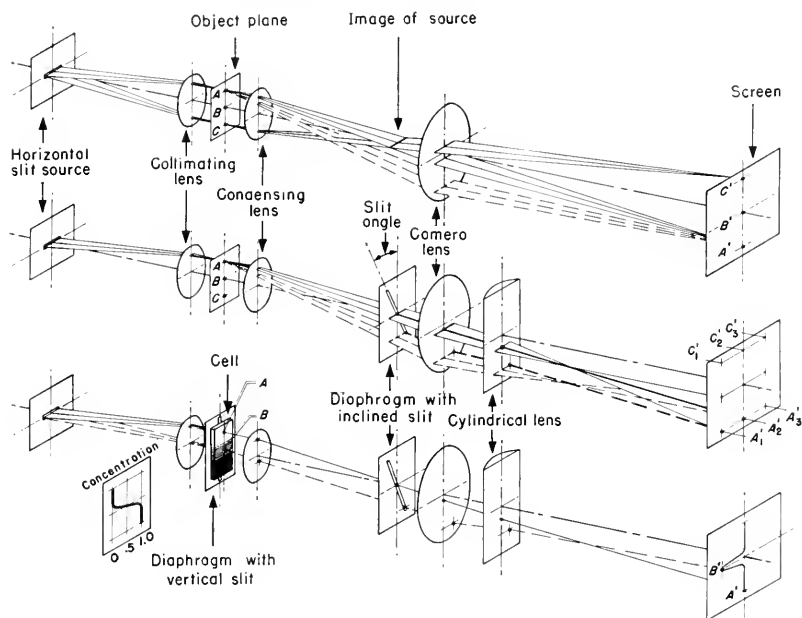
Fig. 6.   Perspective schematic drawing illustrating working principle of cylin-
drical lens method.

are parallel after passing through the collimating lens;   the condens-
ing lens forms an image of the slit source in front of the camera lens
as shown;   sharp images (A', B', C') of the small holes (A, B, C) in
the object diaphragm are formed on the screen by the camera lens
and the condensing lens when the diaphragm is diffusely illuminated.

Limit consideration to those rays that pass through the small holes.
Each set of rays corresponding to one of the holes may be thought of as
making up a "light sheet" that varies in width as it progresses but is always
parallel to the horizontal slit source. If the light sheet diverging from A to-
ward the condensing lens were deviated downward (by a prism, for example)
from the normal course at A, as indicated by the broken lines, it would

nevertheless converge at A' by reason of the fact that the lenses have been positioned to produce such a result. An analogy holds for the other images. Now suppose, as shown in the second arrangement, that a cylindrical lens with vertical axis is added to the system in such a position that a point light source placed at the position of the slit image would be sharply focused as a vertical line image on the screen. The effect of this lens on each of the original light sheets is to converge it prematurely, so that it again diverges and forms a horizontal line image $A'_1 A'_2 A'_3$ on the screen instead of a point image. The vertical positions of the images are not altered by the cylindrical lens. A downward deviation (broken lines) of the sheet at A would not alter the position of the line image. Suppose now that in the vertical plane containing the slit image a thin diaphragm, provided with a narrow inclined slit, is interposed and centered about the optical axis. Then, of the undeviated rays through A, only the central one (heavy unbroken line) is able to pass through the remaining lenses and reach the screen, where it will appear as a point of light A'_2, as contrasted to the previous horizontal line $A'_1 A'_2 A'_3$. Since the undeviated light sheets through A, B, and C converge in a common image of the light source at the second diaphragm, it follows that B and C will also be represented at the screen, when there is no deviation at the object diaphragm, by light points B'_2 and C'_2, which are situated along a central vertical line passing through A'_2. However, if the light sheet from A, for example, is deviated downward (broken lines) sufficiently from its normal course, then only the ray (heavy broken line) at one edge of the light sheet will pass through the remaining lenses, and the light point corresponding to A will appear at A'_1. The horizontal distance by which any light point will be displaced from the central line $C'_2 A'_2$ is directly proportional to the downward displacement suffered by the corresponding light sheet at the object diaphragm. Light coming from B or C is likewise affected. The system is thus able to convert a vertical deviation of light into a horizontal displacement of a light point on a screen (or photographic plate) without altering the point's vertical height, which itself corresponds to a certain level (or certain vertically disposed hole in example given) in the object.

With the elements of the system arranged as described, the light intensity and the resolution at the screen can be increased without affecting the operation of the system by substituting narrow horizontal slits for the holes A, B, and C. In any one light sheet corresponding to one of these slits, more rays will then pass through the inclined slit, diverging from it laterally as they progress toward the camera lens. However, they will be brought to focus again on the screen since the cylindrical lens has been positioned to accomplish this effect. Now suppose, as illustrated in the third arrangement of Figure 6, that a transparent flat-walled cell containing an immobilized sedimentation boundary is substituted for the diaphragm in such a way that the maximum refractive index gradient, *i.e.*, the center of the boundary, is in the

position corresponding to B. The corresponding point B' will be the one suffering the maximum horizontal displacement from the center line on the screen. Light coming through the cell below and above the boundary, as at A, is not deviated and the corresponding light points fall on the center line.

Displacements of light points corresponding to successive positions within the boundary are proportional to the respective refractive index gradients, with the result that the light points form a smooth continuous curve on the screen. With the ultracentrifuge there are extraneous gradients due to distortions of the cell, hydrostatic compression within the contained fluid, and the partial sedimentation of salts or other materials incorporated in the suspending medium. The height and shape of the "base line" curve (with reference to the undeviated light through the counterbalance) vary slightly during the course of
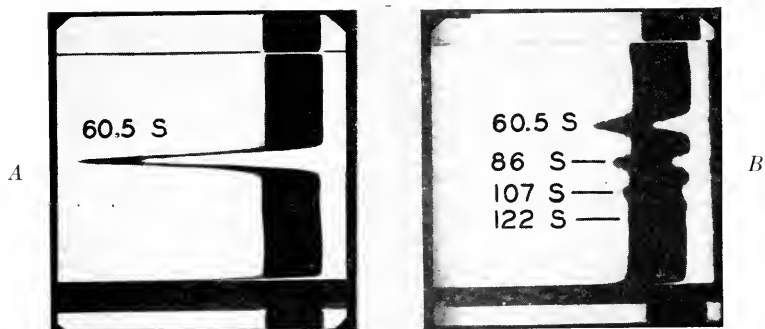
Fig. 7.   Refractive index photographs showing single boundary (A) in a relatively monodisperse preparation of 0.8% hemocyanin, and multiple boundaries (B) produced by irradiating the same material with X rays.

a centrifugation because of the continued redistribution of the sedimentable material within the medium.   Hence, it is necessary to obtain a series of base line photographs during a duplicate run with the suspending medium alone.   The difference (with respect to the undeviated light) between the curves obtained after comparable centrifugation times with and without the principal sedimenting solute represents the effect due to this solute.   The differential curve will be nearly symmetrical if the material is monodisperse and the peak will represent the position of the boundary.   There will be a peak or hump for each differentiated sedimenting component.   The area beneath each peak is proportional to the concentration of the respective component.

In the system utilized in the vacuum ultracentrifuge already described (*89*), an opaque strip or bar is substituted for the inclined slit, so that the deviations of light are indicated by the lateral displacement of a dark band against a bright background. Measurements are taken from both edges of the band and averaged. Examples of sedimentation photographs obtained with such an arrangement may be seen in Figures 5 and 7.

C. MATHEMATICAL THEORY

1. Sedimentation Constant

Consider a monodisperse suspension sedimenting in a centrifuge having an angular velocity of ω radians per second. If at time t, the boundary is at a distance x from the axis of rotation, the sedimentation velocity is dx/dt. The velocity per unit field of force, being a characteristic constant for the sedimenting material, is called the sedimentation constant (*1*) and is denoted by s. Thus:

$$s = \frac{dx}{dt} \frac{1}{\omega^2 x} \tag{1}$$

Sedimentation constants are usually given in S or Svedberg units, *i.e.*, in units of centimeters per second per unit field of force multiplied by 10^{13}. Thus, for hemoglobin $s = 4.4 \times 10^{-13}$ cm./sec./dyne/g. = 4.4S. It is often required to correct a value of the constant from one set of conditions (subscript 1) to some other set (subscript 2). Since sedimentation rate varies directly as the net centrifugal pull on the particle and inversely as the resisting force:

$$s_2 = s_1 \frac{\eta_1(\sigma - \rho_2)}{\eta_2(\sigma - \rho_1)} \tag{2}$$

where η represents viscosity of the medium and σ and ρ represent, respectively, the densities of particle and medium. It is important to note that this factor corrects for changes in η and ρ only and is not sufficient unless it can be assumed that these are the only factors contributing appreciab y to the change in rate. To permit direct comparison of sedimentation constants, this assumption usually is made and the values reduced to standard conditions equivalent to water at 20°C. Thus, if the subscript 20 refers to water at this temperature, the corrected constant (*1*) is:

$$s_{20} = s \frac{\eta(\sigma - \rho_{20})}{\eta_{20}(\sigma - \rho)} \tag{3}$$

2. Frictional Coefficient

In an infinitely dilute solution, the frictional force acting on each mole of sedimenting material per unit sedimentation rate is called the frictional coefficient, f_s. Thus, if the force is proportional to the rate, the resisting force per mole in a centrifugal field is given by:

$$F_s = f_s \frac{dx}{dt} = f_s \, s\omega^2 \, x \tag{4}$$

But F_s must be equal to the net motivating force, F_c, which is simply the difference between their combined centrifugal weight and the buoyancy exerted by the displaced medium. Hence:

$$F_c = M\omega^2 x - \frac{M\rho\omega^2 x}{\sigma} = M\omega^2 x(1 - V\rho) \tag{5}$$

where M is the molecular weight and V is the partial specific volume of the solute, equivalent to $1/\sigma$. From this it follows that:

$$f_s = \frac{M(1 - V\rho)}{s} \tag{6}$$

In the diffusion process, the frictional coefficient f_D for dilute solutions is given by:

$$f_D = RT/D \tag{7}$$

where D is the diffusion constant of the material, T is the absolute temperature, and R is the gas constant (8.313×10^7).

For spherical particles, it is known from the Stokes formula (82) that the frictional coefficient is related to particle size as follows:

$$f_0 = 3\pi d\eta N = 6\pi\eta N \, (3MV/4\pi N)^{1/3} \tag{8}$$

where N is the Avogadro constant and d is the particle diameter.

When spherical particles having a molar frictional coefficient equal to f_0 become enlarged (solvated) symmetrically by combining with some of the suspending medium, the frictional coefficient is increased (83) in proportion to the diameter. Expressed in other terms:

$$f = f_0 \left(\frac{r + V\rho}{V} \right)^{1/3} \tag{9}$$

in which r is the number of grams of the medium combining with one gram of the solute.

An idea of the effect of particle shape on the frictional coefficient can be obtained from the following expressions (84):

For oblong ellipsoids $(k < 1)$:

$$\frac{f}{f_0} = \frac{(1 - k^2)^{1/2}}{k^{2/3} \log \dfrac{1 + (1 - k^2)^{1/2}}{k}} \tag{10}$$

For oblate ellipsoids $(k > 1)$:

$$\frac{f}{f_0} = \frac{(k^2 - 1)^{1/2}}{k^{2/3} \arctan (k^2 - 1)^{1/2}} \tag{11}$$

In these formulas k equals the ratio between the equatorial radius and half the length of the axis of rotation, the particle volume being the same for f and f_0. Also, random orientation of the particles is assumed.

3. Molecular Weight and Particle Size

When conditions are such that $f_s = f_D$, it is possible to equate expressions (6) and (7) and derive the Svedberg formula (1) for molecular weight:

$$M = RTs/D(1 - V\rho) \tag{12}$$

The diameter, d, of spherical particles in dilute preparations may be computed from the following expression, which is obtained by equating the net motivating force and the frictional coefficient as given by equation (8):

$$d^2 = 18\eta s V/(1 - V\rho) \tag{13}$$

For the equilibrium condition, equating expressions for the transport of material by sedimentation and diffusion leads to the following equation (1) for molecular weight in the case of a monodisperse system:

$$M = \frac{2RT \log (c_2/c_1)}{(1 - V\rho)\omega^2(x_2^2 - x_1^2)} \tag{14}$$

where c_1 and c_2 are the concentrations at radial distances x_1 and x_2.

Another form (1) better suited for methods based on the measurement of concentration gradients is:

$$M = \frac{2RT \log [x_1(dc/dx)_2/x_2(dc/dx)_1]}{(1 - V\rho)\omega^2(x_2^2 - x_1^2)} \qquad (15)$$

4. Concentration Correction Factor

The relation between the concentration of a particular component in its respective plateau region at any time t and the original concentration is, as shown by Svedberg and Rinde (20):

$$c_t = c_0(x_o/x_t)^2 \qquad (16)$$

where x_o and x_t are the respective radial distances of the meniscus and the boundary at time t.

D. EXPERIMENTAL REQUIREMENTS

1. Preparation of Material

Under usual conditions most macromolecules are electrically charged with respect to the suspending medium and this circumstance tends to decrease their sedimentation rate, by reason of the attraction and added drag of the more slowly sedimenting ions. The effect can be very large, amounting to more than 20% in many cases. In the equations given in Section C for determining particle weights and sizes, it was assumed that no extraneous forces of this type were active. However, it has been shown that the addition of sufficient low molecular electrolyte is effective in repressing this Donnan effect. In the case of proteins, the reduction in sedimentation rate by the charge of the particle will amount to less than 1% if 0.2 mole of sodium or potassium chloride is added for each 1% protein (85). Also, it is known that pH, the concentration of the protein, temperature, high salt concentrations, and even the presence of another protein may influence the sedimentation constant and these factors must be considered in any ultracentrifugal study. Another precaution often neglected is that of "cleaning up" a preparation by filtration or preferably by differential centrifugation before running it in the ultracentrifuge. Aggregated or extraneous material of high sedimentation rate which will be precipitated before the boundary being studied becomes differentiated can introduce error into the original deter-

minations of particle density, solution viscosity, refractive index increment, and concentration of the principal material.

2. Measurement of Sedimentation Constant

Sedimentation rate is affected very markedly by the concentration of the material. As can be seen from Table I, it decreases more than 10% for each 1% increase in concentration in the case of most serum proteins, even though they are not very asymmetrical. The effect becomes increasingly pronounced with increasing asymmetry of the particles. For the sedimentation rate to have any significance as far as comparison with other values or as far as application to the formulas in Section C is concerned, it is necessary to make determinations with several different concentrations and extrapolate the corrected (for viscosity) results to zero concentration. It is generally desirable to cover at least a threefold range in concentration and to study the material in concentrations at least as low as a few tenths of one per cent. In using equation (12), for example, one cannot assume that s and D are affected to the same degree by any given increase in solute concentration (52). All the equations used in ultracentrifugation studies apply strictly only to infinitely dilute preparations and when it is necessary to apply a viscosity correction, as in equation (2), there is some uncertainty as to whether the viscosity of the medium or the viscosity of the complete preparation should be used. Although the problem is of little consequence when extrapolation to zero concentration is made, it seems preferable to use the complete solution viscosity if only a limited number of experimental determinations can be made (52). In any event, the most important and largest correction generally made is for the change in viscosity brought about by variations in temperature of the specimen. Since this generally amounts to more than 2% per degree Centigrade, it is preferable to know the temperature of the rotor and cell to within a few tenths of one degree. Another factor that must be considered is the increase in viscosity of the medium brought about by the hydrostatic compression. In the case of aqueous solutions in a 1.5 cm. column of fluid at 60,000 r.p.m., the average effect on the sedimentation rate is only about 1%, but with some organic solvents it can be many times this amount (86). In view of other factors (discussed later) that probably compensate for the small viscosity effect in aqueous solutions, it is doubtful whether any correction should be made.

In the derivation of equations (12), (14), and (15) it was assumed that the frictional coefficients in sedimentation and diffusion were equal.   For the equilibrium case, this condition is essentially fulfilled, since the two phenomena are counteracting each other under the same experimental conditions.   For the same reason, it was the early practice to determine the diffusion constant from the progressive spread of the sedimenting boundary in the ultracentrifuge.   This method proved inadequate for several reasons.   First, since sedimentation rate generally increases with increasing dilution, there is a sharpening of the sedimentation boundary as the particles in the trailing edge of the boundary continually tend to overtake the slower moving particles in the plateau region.   (Boundary sharpening can also be caused by slight convective disturbances.)   Secondly, a small percentage of variation in sedimentation rate, as will be caused by slight inhomogeneity or aggregation, can produce sufficient spread or distortion of the boundary's distribution curve to produce a very large error in any determination of the diffusion constant.   Determination of the diffusion constant is now generally made by separate experiments with a regular diffusion cell (87,88).   Obviously, a great deal of uncertainty regarding the equality of the frictional ratios is removed if the ultracentrifugation and diffusion studies are done at nearly the same temperature, in the same medium, and with the same concentrations.   Another question that immediately presents itself is whether there might be a preferential orientation of asymmetrical particles during sedimentation.   There are good theoretical arguments (high thermal compared to centrifugally generated energy) against this for particles in the macromolecular range.   Furthermore, the effect can be tested by centrifuging at different speeds, but no experimental evidence of orientation has been presented.

Still another factor that may affect the frictional coefficient and has a direct bearing on the value of ρ to be used in the equations of Section C is the increase in ρ and σ due to hydrostatic compression in the cell.   In the case of aqueous solutions in the ordinary ultracentrifuge at 60,000 r.p.m., the average increase in the density of the fluid column amounts to about 0.8% (86).   If the sedimenting particle suffers no contraction in size at all, its average rate will be decreased by about 2.5% (assuming $V = 0.75$) because of the increased buoyancy of the medium.   However, if the compressibility ratio of the particle is the same as that of the medium, the rate will increase slightly because of the reduced particle size.   It appears logical that

the latter more nearly represents actuality in most cases and hence it seems unwise, lacking information about the compressibility of the particles, to attempt any correction, especially in view of the fact that the increased viscosity of the compressed medium causes a slight decrease in rate. Neglect of all compressional effects with aqueous solutions could hardly introduce an error of more than 1 or 2% under the conditions cited.

Sedimentation rate is usually determined by photographing sedimenting boundaries at intervals, reducing each measured incremental displacement to the computed amount per unit field of force (basing centrifugal force on the average radial position during the displacement), and averaging all the values obtained. Where multiple boundaries are not resolved until they approach the bottom of the cell, the rate must be based on their effective starting time at the meniscus. An allowance for reflection at the meniscus before clearance of the boundaries can be estimated from runs with monodisperse preparations of similar particle size and concentration. The correction is negligibly small when the total sedimentation time is not more than about three hours. When the position of a diffuse boundary is taken as the point of half concentration, a slight error is introduced into the determination of sedimentation rates (as well as into estimates of diffusion constants) by reason of the fact that diffusion is taking place in a sector-shaped cell. However this error is very small with the large rotors and cells already described and may be neglected (1, p. 22).

3. Determination of Partial Specific Volume

The equations in Section C were derived on the assumption that V, the partial specific volume of the particle, represents the reciprocal of its density under experimental conditions. However, this is seldom known with any exactness because of hydration or solvation of the particle by the suspending medium. This tends to increase both the partial specific volume (unless $\rho > \sigma$) and the frictional coefficient. In case of proteins, the amount of hydration is usually of the order of 20 to 50%. For particles having molecular weights of about 40,000 this may represent an outer shell of water approximately one molecule in thickness. In case of some larger particles, such as viruses there appears to be an actual imbibition (53). Attempts have been made to determine the partial specific volume of hydrated particles by centrifuging them in media of different densities and extrapolating

the results to zero sedimentation. Usually salts, sucrose, urea and other low molecular substances have been used to increase the density of the media but it has been found that these introduce an uncertainty because of their apparent ability to alter the partial specific volume and the frictional ratio by osmotic extraction of water or through partial imbibition of the denser medium (53). Through the use of larger molecules, such as proteins, with a sedimentation rate that is low in comparison to that of the material under study these disadvantages have been largely overcome (61) and the method now appears very promising, particularly for the investigation of viruses.

In applying the equations for molecular weight (equations 1, 14, and 15) it is generally not necessary to know the amount of hydration, as will be shown below. The density of dried or crystallized material can be determined by measuring with a pycnometer the volume displaced by a known weight of substance. If it can be assumed, as is generally the case, that the density of the material being bound to the particle is approximately equal to that of the medium, then the partial specific volume obtained with dried material may be applied to the above equations. In such an instance the value computed for M represents the molecular weight of the particles in such a state that their density is the same as that which characterized the material when the weight of a sample was determined. The reason the density of the unsolvated particle may be used is that the frictional ratio in both sedimentation and diffusion is changed to the same degree in accordance with equation (9) and also that the net centrifugal force acting on the particle has not been appreciably altered in spite of its increased size.

It is important to note that in using equations (2) and (3) one must know the solvated density of the particle or must assume some degree of solvation. Fortunately, correction to the differential density term is seldom very appreciable and probably can be neglected in most cases.

4. Selection of Equipment and Methods

The first requirement of an ultracentrifuge is that its "resolving power" be adequately high. This expresses the ability of an ultracentrifuge to resolve separate components and is also directly related to the possible precision of measurement for a monodisperse system. Svedberg (1, p. 44) has shown that the resolving power is proportional

to $hR\omega^2$ (where h represents the height, measured radially, of the fluid column, R is its average distance from the axis of rotation, and ω is the angular velocity) and has made a comparison of actual values for different machines. For general work with most proteins and other particles of comparable size it is desirable that the resolving power be as high as possible and, as already stated, it has been found that optimum conditions are generally realized with comparatively large rotors operating at speeds in the nieghborhood of 60,000 to 70,000 r.p.m. Limitations on the resolving power and the accuracy of measurement can be appreciated by reference to Figure 1. First, while the boundary is still near the meniscus during an early part of a run, the diffusion process is not a normal one and measurements cannot be started until the boundary begins to clear the meniscus. Also the portion of the cell over which measurements of the boundary's midposition may be continued is quite limited. Some idea of resolving power can be had from the observation that, with the vacuum ultracentrifuge already described spinning at 60,000 r.p.m., it is barely possible to differentiate two equally concentrated "globular" proteins having molecular weights of about 18,000 and 40,000. In such an instance, where resolution of boundaries is not obtained until the faster is approaching the bottom of the cell, it is necessary in the determination of sedimentation rate to know the effective starting time of the boundary at the meniscus. For reasons explained below, this corresponds to the moment of reaching full speed in the oil-driven ultracentrifuge or, in the case of the vacuum ultracentrifuge, to the moment when the rotor has reached about two-thirds of operating speed, assuming a constant rate of acceleration.

A second requirement is that the centrifuge rotor be able to accommodate fluid columns of relatively great thickness (at least 1 cm.) so that measurements can be made with low concentrations. Also adequate means must be provided for measuring the temperature of the fluid to within a few tenths of one degree Centigrade. In the case of the vacuum ultracentrifuge rotor, which is not very responsive thermally to its surroundings and for which the temperature rise during a run is not more than a few degrees, the mean of the starting and final temperatures may be used for purposes of computation (6). In many cases it is desirable that the ultracentrifuge be capable of operation near 0°C. (cell temperature) so as to furnish comparable experimental conditions for those substances whose diffusion constants must be determined at reduced temperatures for best results.

And of course operation of the equipment must be smooth enough to ensure good optical definition in the photographs.

Convective disturbances may be caused by temperature gradients in an ultracentrifuge whenever these gradients have any component that is not radially directed so as to produce a positive density gradient away from the axis of rotation. These disturbances are magnified in proportion to the centrifugal force and with a given temperature gradient are most pronounced with low concentrations of highly diffusible or polydisperse material in large cells (76). These are exactly the conditions that must be attained for precise measurements on particles of low molecular weight, and hence it is of considerable importance to select, especially for such studies, an ultracentrifuge that approaches the convection-free ideal as nearly as possible. It should be noted that a sharp boundary does not indicate an absence of convection; its sharpness may in fact be due partly to one or several circulating currents above or below the boundary, affecting its apparent rate as well as its shape. While viscous drag against the walls of the cell tends to slow down such convection, only sufficient density gradient within the fluid (such as exists within a sharp boundary) can prevent it. Convective disturbances can also be caused by an improper shape or misalignment of the cell (7).

In the oil-driven ultracentrifuges of Svedberg (1) the convection is so pronounced during acceleration of the rotor that no boundary is formed until conditions become uniform at full operating speed. This is apparently due to a rapid and continual rise in rotor temperature because of heating associated with friction from the bearings (which are integral with the rotor and support its full weight), the impinging oil (against turbines), and the enveloping hydrogen. To minimize convection at operating speed, constant and balanced (with respect to the two ends of the rotor) conditions must be maintained. With the vacuum centrifuge convection is minimized by reducing air resistance (pressures of a micron or less) and thermal contact with the bearings as much as possible. In properly designed vacuum ultracentrifuges, no appreciable convection occurs during acceleration. It is therefore important that such machines be capable of very rapid acceleration to full speed in order that a boundary be sedimented away from the meniscus before its concentration distribution becomes unduly affected by "reflection" from the meniscus. This is especially important when the starting position of the boundary is used in the determination of sedimentation rate. It is of interest in this connec-

tion that a Duralumin rotor can be accelerated almost three times as
rapidly as a steel one because of its correspondingly lower specific
gravity.   Also, at full speed it requires proportionately less barricad-
ing to furnish adequate protection to operator and equipment in the
event of rotor failure.   And besides presenting a considerably simpler
machining problem, it has the very great advantage of high conduc-
tivity and hence low susceptibility to temperature gradients and con-
vective disturbances.

   Selection of an optical system generally resolves itself into a con-
sideration of the various refractive index methods, since they in gen-
eral offer distinct advantages over the absorption method, which
requires knowledge, not too simply or precisely acquired, of the rela-
tionship at every level in the cell between concentration and the cor-
responding amount of photographic blackening in the image of the
revolving cell.   As for the refractive methods, the Lamm method
using a real scale is undoubtedly subject to the fewest optical errors.

   If an optically projected scale is used, the zonal errors of the pro-
jection system must be negligible or else accurately measured and ap-
plied as corrections.   The cylindrical lens method involves the great-
est number of optical parts, each of which must be precisely posi-
tioned and oriented, and is hence most susceptible to error if im-
properly designed or assembled.   It is essential that the lenses be of
good optical quality.   Apparently most of the disappointments with
such systems have been caused by improper assembly or a lack of
accuracy in the cylindrical lens.   Because of the limited demand for
precise cylindrical lenses, there are few optical concerns that have
the specialized experience and equipment required for grinding them
accurately.   If light is paralleled through the cell (to minimize paral-
lax errors) and if good lenses are used, focal lengths can be relatively
short (of the order of 50 cm. for the camera lens) and the total length
of the optical system can be made as little as 3 m.   With any well
designed system of the refractive index type the limiting accuracy is
determined primarily by the optical resolution of the system, and for
measuring the sedimentation rates of distinct boundaries, a good
cylindrical lens system appears to be comparable in accuracy to the
scale method, although the latter may afford slightly greater accuracy
in the determination of the actual distribution of material within a
boundary (25).   In general, the very convenient cylindrical lens
method can be made quite adequate for sedimentation velocity stud-
ies, whereas use of the scale method may be warranted for equilibrium
determinations.

When it is desired to apply the equilibrium method to a particular preparation, it should be first studied by the sedimentation velocity method if possible to ascertain the degree of ultracentrifugal homogeneity. Unless the material is relatively monodisperse, it is not ideally suited for study by the equilibrium method, although an "average" value for the molecular weight can be obtained even with mixtures (1). Columns of fluid are generally made shorter than for velocity measurements in order to shorten the time required to attain equilibrium, which is proportional to the square of the cell height (radially) and inversely proportional to the diffusion constant (1, p. 56). For example, even with a fluid column only 5 mm. tall, almost four days are required for a protein having a molecular weight of 40,000 and a diffusion constant of 8×10^{-7}.

E. INTERPRETATION OF RESULTS

1. Significance of Frictional Ratio

The frictional ratio f/f_0 can be obtained for a substance of known molecular weight and sedimentation constant by comparing frictional coefficients computed according to equations (6) and (8). It is essentially the ratio between the actual frictional coefficient of a substance in solution and the value it would have if the particles were spherical and had an effective density in solution equal to the value ($\sigma = 1/V$) used in the equations. The only density value generally known is that for the dried, unsolvated material, and hence the departure of the ratio from unity may be due to either or both solvation and asymmetry of form. Values for some of the serum proteins are given in Table I. From what is known about proteins, there is, in accordance with equation (9), some assurance that hydration alone would seldom account for an increase in the ratio by a factor of more than about 1.2 or 1.25. By dividing the experimentally determined ratio by the assumed ratio due to solvation, one obtains a new ratio that may be used to gain some idea of particle shape through the application of equations (10) and (11). For example, ratios of 1.10, 1.20, and 2 applied to equation (10) yield length to diameter ratios of 2.9, 4.4, and 20, respectively.

2. Determination of Homogeneity

Ultracentrifugal homogeneity of a boundary can generally be judged by comparing its shape and rate of spreading to values ob-

TABLE I

Molecular Constants of Some Serum Proteins[a]

Protein	S_{20}	D_{20}	V_{20}	Mol. wt.	f/f_0
Fetuin					
Calf.........	$3.28-0.0020 \times \Delta n^b$	5.5	0.71_4	50,600	1.60
Cow fetus....	$3.09-0.0015 \times \Delta n$	5.0	0.69_2	48,700	1.80
Cow fetus....	$3.23-0.0020 \times \Delta n$	(5.5)	0.71_2	49,500	1.61
Serum albumin					
Human......	$4.6-0.0022 \times \Delta n$	5.9	0.73_6	72,300	1.30
Cow fetus....	$4.7-0.0019 \times \Delta n$	6.4	(0.736)	67,800	1.21
Globulin					
Cow γ-globulin					
$P_{II}\delta_1$......	$7.4-0.0040 \times \Delta n$	3.7	(0.732)	180,000	1.52
$g_1\delta_7$........	$7.6-0.0025 \times \Delta n$	3.5	0.73_2	192,000	1.56
$g_1\delta_8$........	$7.6-0.0029 \times \Delta n$	3.8	(0.732)	182,000	1.49
Human γ-globulin......	$7.1-0.0034 \times \Delta n$	4.0	0.71_8	153,000	1.51
Anticomplementary human pseudoglobulin....	7.1	3.8	0.73_2	170,000	1.53
20-components					
Isohemagglutinin, human	$19.8-0.0074 \times \Delta n$	...	(0.73)	($\sim5 \times 10^4$)	..
Pathological human euglobulin					
From β_7....	$19.1-0.022 \times \Delta n$	>1.4}	0.73_3	$\sim1.2 \times 10^6$	2.1
From β_9...	$19.5-0.022 \times \Delta n$	<1.6}	(0.733)	$\sim11.2 \times 10^6$	2.0
X-protein					
Human......	5.9	1.7	$0.96_9{}^c$	($\sim2 \times 10^6$)^c	..

[a] After Pedersen (28). S_{20} values are in Svedberg units and D_{20} values are in units of 1×10^{-7}.

[b] For most serum proteins, Δn equals approximately 189 times the concentration of protein in per cent.

[c] For the hydrated particle.

tained by separate diffusion experiments.   There is an increase in boundary sharpness with increasing concentration in the case of ultracentrifugation and, for the most precise work, results must be extrapolated to zero concentration or some allowance must be made. With experience, one learns to recognize immediately by inspection of the photographs any appreciable inhomogeneity.   The sharpening with concentration is sometimes actually an advantage, in that it permits greater resolution of components to be obtained.   The amount

of spreading due to diffusion alone can also be ascertained, at least
with a vacuum ultracentrifuge, by decelerating to a very low speed
before the boundary reaches the bottom of the cell and allowing it to
diffuse, with negligible sedimentation, at this low speed (see Fig. 7).
With large particles, such as viruses especially, aggregation appears
to be a common phenomenon (54) and may produce either well de-
fined multiple boundaries or a single blurred boundary with the
trailing edge having a rate characteristic of the primary particles.
For the first photograph (A) in Figure 7, the speed of the ultracen-
trifuge was reduced to a very low value, after the required boundary
displacement had been accomplished, in order to minimize boundary
sharpening that occurs during rapid sedimentation and to permit
free diffusion of the material for a determination of the diffusion con-
stant.   Displacement of base line toward the left in the second picture
(B) is due principally to hydrostatic compression of the fluid by the
centrifugal force.

A boundary may be made more or less sharp and its rate affected
if the primary particles are in equilibrium with dissociation or as-
sociation products whose relative concentration depends on the con-
centration of the primary particles (1, p. 28)   Many proteins appear
to suffer partial dissociation in concentrations much below 0.5% and
some association in concentrations much above 1%.   This limits the
range that can be employed usefully for extrapolating the results to
zero concentration.

3. Accuracy and Limitations of the Method

With an ideal material (high molecular weight and stability) de-
terminations of sedimentation rate can be repeated with an average
deviation of the order of 1% or better when concentrations of at
least a few tenths of one per cent are used and the optimum condi-
tions realizable in actual practice prevail.   In most cases the greatest
error probably occurs in the measurement of cell temperature.   At-
tainable accuracy necessarily decreases as more diffusible or less
stable material is investigated.   In a typical study of the smaller
serum proteins, for example, an average deviation of several per cent
is not uncommon and an extreme difference between individual de-
terminations of as much as 10% is not unexpected.   If a sedimenta-
tion constant is to be determined with any degree of precision, at
least several different determinations must be made.   Even when a
mean value is based on many careful determinations made with a

single ultracentrifuge, it is very doubtful whether the value can be regarded as significant to within less than 1 or 2% in most cases.

Boundaries can be detected and their rates measured (with decreasing accuracy) with concentrations as low as about 0.01% in the case of the larger monodisperse proteins. For molecular weights below several hundred thousand, several hundredths of one per cent may be required. The concentrations of fully resolved components can usually be determined from the area under the refractive index curves to within a few per cent when the value is of the order of 1% or more, the accuracy depending on the quality of the optical system and the spread of the boundary. Even when boundaries are not completely separated, their respective concentrations can be estimated by a method of interpolation (1).

It is not possible to make any general statement regarding the accuracy of a molecular weight determination by either the velocity method or the equilibrium method, since so many variables are involved. Discrepancies of at least a few per cent are the general rule and one is probably not justified in considering the usual determination significant to within less than 5 to 10%.

4. Representative Applications

Thousands of practical applications of the ultracentrifugal method have been reported in the literature and bibliographies covering many of these may be found in the publications of Svedberg and associates (1,5). Less extensive bibliographies have been given by Pickels (2). A few typical publications have been listed in the bibliography at the end of this article. The greatest advantage of the method seems to have been realized in the study of proteins and viruses. Table I summarizes representative findings for some of the serum proteins.

Of interest in the way of pointing out some of the problems that may be encountered as well as some of the avenues of investigation that may be explored, are descriptions of the following:

Splitting of hemocyanin molecules by ultrasonic waves (32); dissociation of hemocyanin (33); stability in protein systems (51); catalytic effect of papain on thyroglobulin (37); interaction of proteins (38); molecular association of hemocyanin produced by X rays (40); effect of ultraviolet radiation and X rays on serum albumin (43); shape of polystyrene molecules (45); splitting of protein molecules by ultraviolet light and X rays (49); influence of pH on stability of equine encephalomyelitis virus (57); effects of

desiccation (*29*); formation of micelles by detergents (*30*); ultracentrifugal homogeneity of carbohydrates (*31*); ultracentrifugal behavior of cellulose (*50*).

F. PREPARATIVE CENTRIFUGATION AND QUANTITA-TIVE METHODS BASED ON SAMPLING

1. Preparative Centrifugation

Most high speed (above 10,000 r.p.m.) preparative centrifuges suitable for the concentration and purification of particles in the size

Fig. 8.   Six-inch preparative rotor of Duralumin for routine operation at 50,000 r.p.m. in the vacuum ultracentrifuge.   (Courtesy *Specialized Instruments Corporation*.)

range below 100 μ are of either the inclined tube type (*62,63,76*) or the continuous flow type (*91*), and are operated in a vacuum or under atmospheric conditions depending on the required speed, capacity, operating temperature, and degree of freedom from convective disturbances.   One of the interchangeable preparative rotors used with the vacuum ultracentrifuge already described is shown in Figure 8. In such rotors the tube holes are usually arranged at some angle between 10 and 45° to the axis of rotation, and the rotor is machined

from a solid piece of metal (Duralumin) for reasons of strength. For well designed rotors capable of accommodating between 100 and 400 ml. of fluid, the maximum rotational speed and the maximum force on the fluid are approximately proportional to the reciprocal of the rotor diameter (62). With a properly constructed rotor 6 in. in diameter, one can attain routine operating speeds up to about 60,000 r.p.m. and forces above 200,000 times gravity (62). In such fields of force all viruses can be separated readily and even particles in the size range of the serum proteins can be sedimented, though not with so sharp a differentiation. Tubes for containing the fluid are usually thin-walled and of either Celluloid or metal. In the inclined tubes, particles (except very large ones of high sedimentation rate) do not pack against the walls and slide to the bottom of the tube as a precipitate by reason of their diffusibility. Instead, there is a continual circulation of fluid caused by the downward movement of the more concentrated layer of fluid continually being formed by sedimentation against the outer wall, and the actual precipitation of particles occurs near the bottom of the tube (76). Most rapid separation and the greatest capacity of fluid are possible when the inclination of the tubes is made small (10 to 20°). In Figure 8 are ten Celluloid tubes inclined at a 26° angle to the axis and having a total capacity of 130 ml. Tubes are individually sealed by expansion of rubber plugs in caps and the rotor itself is hermetically sealed with the aid of two rubber gaskets. A knurled rod, internally threaded at one end, is used for extracting capped tubes.

High speed angle centrifuges are commercially available (89,93, 94) in a variety of forms. For further discussion of the general field of preparative centrifugation, the reader is referred to a previous publication (2). Also, a list of pertinent references (with parenthetical notes) has been included in the bibliography at the end of the present article. These deal with design, typical applications, and the problems connected with preparative centrifugation. Commercially available centrifuges (91) of the continuous flow type that operate in the open air have been found practical for the large scale concentration of some of the larger viruses, and the reader is referred to the original papers by McIntosh and Selbie (73), Stanley (74), and Taylor (75) for discussions of the efficiency of such centrifuges. These machines are operated by compressed air at speeds up to about 50,000 r.p.m. and the rotating member through which the fluid passes is essentially a hollow tube a few inches in diameter.

2. Sampling Techniques

It is often necessary or advisable to establish a relationship between the material forming an optically recorded sedimentation boundary and some biological or chemical activity. To relate the distribution of some specific property with the boundary, Tiselius, Pedersen, and Svedberg (77) have provided the optical cell of the ultracentrifuge with a porous partition near its center. After a boundary is observed to have passed through this filter, the machine is stopped and the fluid above and below the partition collected for analysis. The angle centrifuge has also been employed with advantage for the measurement of approximate sedimentation rates based on sampling (76,78). With proper technique (79) fluid columns can be sampled with fair accuracy at any number of levels. Advantages of the method are that comparatively large samples may be had for analysis and that it affords a knowledge of the concentration of a specific property throughout the fluid column. Relatively sharp sedimentation boundaries can be obtained and approximate homogeneity as well as average rate determined. Dilute preparations that would otherwise be remixed by convection can be stabilized with a graded concentration of sucrose or some other nonsedimenting material (76).

Centrifuges making use of the "inverted capillary" technique (80) have also been used with good success in the study of several viruses. The material to be investigated is placed in small capillaries and the sedimenting particles are allowed to migrate into a larger reservoir filled with the same material. After centrifugation, the degree of sedimentation is estimated by comparing the average concentrations of activity in the capillaries and the reservoir. Of interest also are the opaque, air-driven ultracentrifuges of the "spinning top" variety, possible applications of which are exemplified by the study of Ruben and associates (81) with radioactively tagged molecules.

References

GENERAL

1. Svedberg, T., and K. O. Pedersen, *The Ultracentrifuge*. Clarendon Press Oxford, 1940 (most complete reference; all aspects of analytical ultracentrifugation, with extensive bibliography).

2. Pickels, E. G., "High-Speed Centrifugation," in J. Alexander, *Colloid Chemistry*, Vol. V, Reinhold, New York, 1944 (analytical and preparative centrifuges; design, theory, application; bibliography).

3. MacInnes, D. A., W. J. Archibald, J. W. Beams, W. B. Bridgman, A. Rothen, and J. W. Williams, *Ann. N. Y. Acad. Sci.*, **43**, 173 (1942) (conference on ultracentrifuge).

4. Pickels, E. G., *Chem. Revs.*, **30**, 341 (1942) (general discussion; optical system).

5. Svedberg, T., *Chem. Revs.*, **20**, 81 (1937) (principles and applications of oil-driven ultracentrifuge).

6. Bauer, J. H., and E. G. Pickels, *J. Exptl. Med.*, **65**, 565 (1937) (air-driven, air-supported ultracentrifuge).

7. Pickels, E. G., *Rev. Sci. Instruments*, **13**, 426 (1942) (cell design).

8. Beams, J. W., *J. Wash. Acad. Sci.*, **37**, 221 (1947) (magnetic and air support, electrical and air drives).

9. Svedberg, T., and J. B. Nichols, *J. Am. Chem. Soc.*, **45**, 2910 (1923) (first optical ultracentrifuge).

10. Skarstrom, C., and J. W. Beams, *Rev. Sci. Instruments*, **11**, 398 (1940) (electrical drive, magnetic support).

11. Pickels, E. G., *Rev. Sci. Instruments*, **9**, 358 (1938) ("turret" type air drive and support bearing).

12. Svedberg, T., and A. Lysholm, *Nova Acta Regiae Soc. Sci. Upsaliensis*, vol. extra. ord. (1927) (first oil-driven ultracentrifuge).

13. Boestad, G., K. O. Pedersen, and T. Svedberg, *Rev. Sci. Instruments*, **9**, 346 (1938) (modern oil-driven ultracentrifuge).

14. Svedberg, T., and B. Sjögren, *J. Am. Chem. Soc.*, **51**, 3594 (1929) (equilibrium centrifuge).

15. Henriot, E., and E. Huguenard, *J. phys. radium*, **8**, 433 (1927) (original air-driven "top").

16. Beams, J. W., E. G. Pickels, and A. J. Weed, *J. Chem. Phys.*, **2**, 143 (1934) (optical centrifuge of "spinning top" type).

17. Beams, J. W., and E. G. Pickels, *Rev. Sci. Instruments*, **6**, 299 (1935) ("spinning tops" and early vacuum centrifuges).

18. Stern, K. G., *Science*, **95**, 561 (1942) (small plastic ultracentrifuge).

19a. Pickels, E. G., and J. W. Beams, *Science*, **81**, 342 (1935).

19b. McBain, J. W., and A. H. Lewis, *J. Phys. Chem.*, **43**, 1197 (1939); *Science*, **89**, 611 (1939) (spinning top).

19c. Nichols, J. B., and E. D. Bailey, "Determinations with the ultracentrifuge," in *Physical Methods of Organic Chemistry*, 2nd ed. A. Weissberger, ed. Interscience, New York, 1949, chap. XIII.

OPTICAL METHODS

20. Svedberg, T., and H. Rinde, *J. Am. Chem. Soc.*, **46**, 2677 (1924) (light absorption method).

21. Philpot, J. St. L., *Nature*, **141**, 283 (1938) (cylindrical lens method).

22. Andersson, K. J. I., *Nature*, **143**, 720 (1939) (cylindrical lens method).

23. Svensson, H., *Kolloid-Z.*, **90**, 141 (1940) (cylindrical lens method).

24. Lamm, O., *Nova Acta Regiae Soc. Sci. Upsaliensis*, **4,** 10 (1937) (scale method).

25. Kegeles, G., and L. J. Gosting, *J. Am. Chem. Soc.*, **69,** 2516 (1917) (interference method for diffusion).

26. Svensson, H., *Nature*, **161,** 234 (1948) (scale method with shortened optical system).

REPRESENTATIVE STUDIES OF PROTEINS, CARBOHYDRATES, ETC.

27. Pedersen, K. O., *Z. physik. Chem.*, **A170,** 41 (1934) (ultracentrifugation of inorganic salts).

28. Pedersen, K. O., *Ultracentrifugal Studies on Serum and Serum Fractions.* Almqvist and Wiksells, Upsala, 1945 (serum proteins).

29. Brosteaux, J., and I. B. Eriksson-Quensel, *Arch. phys. biol.*, **12,** No. 4 (1935) (effects of desiccation).

30. Smith, E. L., and E. G. Pickels, *Proc. Natl. Acad. Sci. U. S.*, **26,** 272 (1940) (detergents).

31. Svedberg, T., and N. Gralén, *Nature*, **142,** 261 (1938) (carbohydrates).

32. Brohult, S., *Nature*, **140,** 805 (1937) (splitting of hemocyanin).

33. Brohult, S., and S. Claesson, *Nature*, **144,** 111 (1939) (dissociation of hemocyanin).

34. Brohult, S., *Nova Acta Regiae Soc. Sci. Upsaliensis*, **12,** 1 (1940) (hemocyanin).

35. Deutsch, H. F., *J. Biol. Chem.*, **169,** 437 (1947) (whey proteins).

36. Gralén, N., *J. Colloid Sci.*, **1,** 453 (1946) (lignin).

37. Lundgren, H. P., *J. Biol. Chem.*, **138,** 293 (1941) (effect of papain on thyroglobulin).

38. Pedersen, K. O., *Proc. Roy. Soc. London*, **B127,** 20 (1939) (interaction of proteins).

39. Pedersen, K. O., *J. Phys. & Colloid Chem.*, **51,** 164 (1947) (fetuin).

40. Pickels, E. G., and R. S. Anderson, *J. Gen. Physiol.*, **30,** 83 (1946) (effect of X rays on hemocyanin).

41. Rothen, A., *J. Gen. Physiol.*, **24,** 203 (1940) (ribonuclease).

42. Rothen, A., *J. Biol. Chem.*, **152,** 679 (1944) (ferritin).

43. Sanigar, E. B., L. E. Krejci, and E. O. Kraemer, *Biochem. J.*, **33,** 1 (1939) (effect of X rays on serum albumin).

44. Sharp, D. G., A. R. Taylor, H. Finkelstein, and J. W. Beard, *Proc. Soc. Exptl. Biol. Med.*, **42,** 459 (1939) (normal tissue components).

45. Signer, R., *Trans. Faraday Soc.*, **32,** 296 (1936) (polystyrenes).

46. Smith, E. L., *J. Biol. Chem.*, **164,** 345 (1946) (immune globulins).

47. Smith, E. L., *J. Biol. Chem.*, **165,** 665 (1946) (immune lactoglobulins from whey).

48. Stern, K. G., and R. W. G. Wyckoff, *J. Biol. Chem.*, **124,** 573 (1938) (catalase).

49. Svedberg, T., and S. Brohult, *Nature*, **143**, 938 (1939) (effect of X rays and ultraviolet on proteins).
50. Jullander, I., *J. Polymer Sci.*, **2**, 329 (1947) (cellulose).
51. Lundgren, H. P., and J. W. Williams, *J. Phys. Chem.*, **43**, 989 (1939) (stability in protein systems).

ANALYTICAL INVESTIGATION OF VIRUSES

52. Lauffer, M. A., *J. Am. Chem. Soc.*, **66**, 1188 (1944) (tobacco mosaic).
53. Smadel, J. E., E. G. Pickels, and T. Shedlovsky, *J. Exptl. Med.*, **68**, 607 (1938) (vaccinia).
54. Pickels, E. G., and J. E. Smadel, *J. Exptl. Med.*, **68**, 583 (1938) (vaccinia).
55. Beard, J. W., D. G. Sharp, A. R. Taylor, I. W. McLean, Jr., D. Beard, A. E. Feller, and J. H. Dingle, *Southern Med. J.*, **37**, 313 (1944) (influenza).
56. Sharp, D. G., A. R. Taylor, and J. W. Beard, *J. Biol. Chem.*, **163**, 289 (1946) (papilloma).
57. Taylor, A. R., D. G. Sharp, and J. W. Beard, *J. Infectious Diseases*, **67**, 59 (1940) (equine encephalomyelitis).
58. Lauffer, M. A., *J. Biol. Chem.*, **143**, 99 (1942) (bushy stunt).
59. Knight, C. A., and W. M. Stanley, *J. Biol. Chem.*, **141**, 29 (1941) (cucumber)
60. Pickels, E. G., and J. H. Bauer, *J. Exptl. Med.*, **71**, 703 (1940) (yellow fever)
61. Sharp, D. G., A. R. Taylor, I. W. McLean, Jr., D. Beard, and J. W. Beard, *Science*, **100**, 151 (1944) (density determination of viruses).

PREPARATIVE CENTRIFUGATION

62. Pickels, E. G., *Rev. Sci. Instruments*, **13**, 101 (1942) (design of angle centrifuges).
63. Masket, A. V., *Rev. Sci. Instruments*, **12**, 277 (1941) (individually plugged tubes, angle centrifuge).
64. Taylor, A. R., D. G. Sharp, D. Beard, and J. W. Beard, *J. Infectious Diseases*, **72**, 31 (1843) (equine encephalomyelitis).
65. Beard, J. W., W. R. Bryan, and R. W. G. Wyckoff, *J. Infectious Diseases*, **65**, 43 (1939) (rabbit papilloma virus).
66. Cunha, R., M. L. Weil, D. Beard, A. R. Taylor, D. G. Sharp, and J. W. Beard, *J. Immunol.*, **55**, 69 (1947) (Newcastle disease virus).
67. Harris, T. N., *J. Exptl. Med.*, **87**, 41 (1948) (hemolytic streptococcus).
68. Kabat, E. A., M. Heidelberger, and A. E. Bezer, *J. Biol. Chem.*, **168**, 629 (1947) (ricin).
69. Lauffer, M. A., and G. L. Miller, *J. Exptl. Med.*, **80**, 525 (1944) (influenza virus).

70. McKinstry, D. W., H. M. Terrill, E. H. Reading, and K. V. Huffman,
       J. Franklin Inst., **238**, 310 (1944) (poliomyelitis virus).
71. Taylor, A. R., D. G. Sharp, H. Finkelstein, D. Beard, and J. W. Beard,
       Proc. Soc. Exptl. Biol. Med., **43**, 648 (1940) (normal tissue compo-
       nents).
72. Taylor, A. R., D. G. Sharp, I. W. McLean, Jr., D. Beard, J. W. Beard,
       J. H. Dingle, and A. E. Feller, *J. Immunol.*, **48**, 361 (1944) (swine in-
       fluenza virus).
73. McIntosh, J., and F. R. Selbie, *Brit. J. Exptl. Pathol.*, **21**, 153 (1940)
       (adaptation of continuous flow centrifuge for virus concentration).
74. Stanley, W. M., *J. Immunol.*, **53**, 179 (1946) (efficiency of continuous
       flow centrifuges).
75. Taylor, A. R., *J. Biol. Chem.*, **163**, 283 (1946) (papilloma virus in continu-
       ous flow centrifuge).

QUANTITATIVE METHODS BASED ON SAMPLING

76. Pickels, E. G., *J. Gen. Physiol.*, **26**, 341 (1943) (study of sedimentation
       in angle centrifuges).
77. Tiselius, A., K. Pedersen, and T. Svedberg, *Nature*, **140**, 848 (1937)
       (separation cell for analytical ultracentrifuge).
78. Curnen, E. C., E. G. Pickels, and F. L. Horsfall, *J. Exptl. Med.*, **85**,
       23 (1947) (pneumonia virus in angle centrifuge).
79. Hughes, T. P., E. G. Pickels, and F. L. Horsfall, *J. Exptl. Med.*, **67**,
       941 (1938) (sampling technique, angle centrifuge).
80. Elford, W. J., and I. A. Galloway, *Brit. J. Exptl. Pathol.*, **18**, 155 (1937)
       ("inverted capillary" method).
81. Ruben, S., M. D. Kamen, and L. H. Perry, *J. Am. Chem. Soc.*, **62**, 3450
       (1940) (application of "spinning top").

MISCELLANEOUS

82. Stokes, G. G., *Mathematical and Physical Papers*, Vol. III, Cambridge
       Univ. Press, Cambridge, 1901, pp. 59–60 (Stokes formula for spherical
       particles).
83. Kraemer, E. O., *J. Franklin Inst.*, **229**, 393 (1940) (effect of solvation).
84. Herzog, R. O., R. Illig, and H. Kudar, *Z. physik. Chem.*, **A167**, 329
       (1933) (effect of particle shape).
85. Tiselius, A., *Kolloid-Z.*, **59**, 306 (1932) (suppression of Donnan effect).
86. Mosimann, H., and R. Signer, *Helv. Chim. Acta*, **27**, 1123 (1944) (effect
       of hydrostatic pressure on viscosity and density).
87. Claesson, S., *Nature*, **158**, 834 (1946) (diffusion cell).
88. Neurath, H., *Chem. Revs.*, **30**, 357 (1942) (diffusion constant determina-
       tions).

COMMERCIAL SOURCES

89. Specialized Instruments Corporation, Belmont, Calif.
90. LKB Produkter Fabriks AB, Stockholm, Sweden.
91. Sharples Corporation, Philadelphia, Pa.
92. Fisher Scientific Co., Pittsburgh, Pa.
93. International Equipment Co., Boston, Mass.
94. Ivan Sorvall, Inc., 210 Fifth Ave., New York, N. Y.
95. Gamma Scientific Co., Great Neck, L. I., N. Y.

VISCOSITY MEASUREMENTS

L. V. HEILBRUNN, *University of Pennsylvania*

A. INTRODUCTION

There are two chief reasons that a biologist or a biophysicist might be interested in viscosity measurement. He might want to know something about the condition of flow of the circulating fluids of an organism. He would then seek to determine the viscosity of the blood or the lymph. Such determination can be made rather simply with ordinary physical methods not peculiar to the student of biological phenomena. A much more difficult problem is the determination of the viscosity of the protoplasm within the cells of an organism. The fact that protoplasm is a colloid and as such is subject to sudden and sharp changes in viscosity makes it obvious that the student of cell mechanics should acquire what information he can about the viscosity of the protoplasm and how this viscosity changes during the life and death of the cell. No other physical property of the protoplasm offers as much hope for the interpretation of vital mechanics as does the viscosity. A cell may become violently active or may lose its activity; it may become poisoned, electrocuted, or crushed and yet such physical properties as specific gravity, electric potential at the surface, or birefringence may scarcely be affected. Not so the viscosity. A dividing cell goes through an entire cycle of protoplasmic

viscosity changes and, during division, the viscosity may jump to four or five times its original value and back again. When muscle and nerve are thrown into activity, the protoplasmic viscosity almost certainly changes, and when a cell is subjected to an anesthetic like ether, the viscosity may be profoundly affected. Indeed, at the present time, the only plausible theory of stimulation and response is a colloidal theory that involves the assumption of marked viscosity change within the protoplasm. Cells are aroused to activity by various diverse agents such as electric shock, ultraviolet radiation, sudden pressure, etc. All these agents could scarcely induce any one type of chemical reaction, nor could they all directly affect the speed of protoplasmic reactions in any definite way. The fact that the chemical reactions of cells are affected in much the same manner by widely different physical and chemical agents can best be explained by assuming that these varied agents cause some universal type of colloidal change and that this change influences the reactions, throws them into gear, as it were. Moreover, if one is to explain physical phenomena like the contraction of muscle protoplasm, one must seek an interpretation in physical terms. No matter how many interesting and complicated chemical reactions may be described in muscle cells or extracts made from them, somehow or other these reactions must be related to physical change in the protoplasmic colloid, either as a result or a cause. And if we are to study physical change in the protoplasmic colloid, viscosity is the best index of such changes that has yet been discovered.

A point that cannot be emphasized too strongly is that viscosity change of the protoplasmic colloid must be studied directly. It is true that the protoplasm is made up of proteins, but the colloidal behavior of the protoplasm is not like that of any known protein. The very fact that the protoplasm is alive indicates that it must have peculiarities of colloidal behavior, and, indeed, it has been abundantly proved that in many important ways the protoplasmic colloid behaves unlike typical nonliving proteins (*cf. 1*). Nor is it permissible to extract pure proteins from a tissue like muscle and then to assume that the behavior of these purified proteins outside the cell is like that which they would have within the living muscle. Thus, the fundamental basis of much of Szent-Györgyi's recent work on muscle is wrong, as can be clearly demonstrated by studies on living, intact muscle cells (*2*).

In the ordinary study of colloid chemistry, major changes in vis-

cosity such as are involved in the change from sol to gel and back again can readily be observed. During gelation, the colloidal fluid in a test tube or beaker suddenly stiffens and, as can be seen at a glance, it will no longer flow when the tube or beaker is tilted or turned upside down. But in a cell the protoplasm may change from a very fluid to a very solid state, and yet, as far as one can see with the microscope, there is no visible indication of any change whatsoever.

Unfortunately, the literature on protoplasmic viscosity has been badly confused by the reckless and careless statements of workers who have not used standard methods of viscosity determination. It is a simple matter to poke a cell with a microneedle, but, from such poking, it is not possible even to obtain a correct estimate of the order of magnitude of protoplasmic viscosity. With a microdissection apparatus, a cell may be attacked with a fine glass needle. Ordinarily, the distensible cell membrane yields to the pressure of the needle so that there is an indentation of the needle into the cell. It is very difficult to determine whether the needle actually enters the protoplasm or not. If it does enter, it almost certainly causes pronounced injury in whatever region it reaches. The needle is of course controlled mechanically, so that one has no sense of the force involved in moving the needle about. Under such conditions the "measurement" of viscosity is mere guess work.

B. VISCOSITY CONCEPT

People generally distinguish liquids that are thick and heavy from those that are thin or light. The distinction is made not on the basis of density but rather of the ease with which a liquid flows or pours, or on the basis of how easily it may be stirred. Scientifically, liquids that flow or pour easily and offer relatively little resistance to stirring are less viscous or more fluid than those that flow with difficulty and offer marked resistance to a stirrer. As a matter of fact the terms viscosity and fluidity are rather well established in common language. The exact definition of viscosity is not so easy to present. The definition commonly employed is that formulated by Maxwell. A convenient wording of this definition is quoted by Barr (3):

"The coefficient of viscosity of a fluid is the numerical value of the tangential force on unit area of either of two parallel planes at unit distance apart when the space between these planes is filled with the fluid in question and one of the planes moves with unit velocity in its own plane relatively to the other."

If the moving plane were to travel at two or three times unit velocity, the force exerted on the stationary plane ought to increase two or three times, and this would be true if no elastic forces were involved.   But, if there is an elastic material between the two planes, as the moving plane traveled at greater and greater speed, part of the extra force it exerted would be expended in overcoming the elastic forces of the material between the two planes, and the force exerted on the stationary plane would not increase proportionately.   In such a system, as the moving plane traveled more rapidly and developed a greater shearing force, the measured viscosity would be less.   Such a condition is met with in plastic solids such as clay or wax and is also found in many colloidal solutions.   For the case of the plastic solid, Bingham (4) proposed the term plasticity, and this term has had wide usage.   However, as Barr (3) points out, Bingham does not speak of plastic liquids, and he reserves the term for solids not continuously deformable by a small shearing force   It is now customary to use the terms "apparent viscosity" or "anomalous viscosity" for those cases in which the viscosity varies with the shearing force.

For further discussion of the viscosity concept, see other published works (5-8, 33).

C. VISCOSITY OF BLOOD

Ordinary methods of viscometry can readily be used in the study of the viscosity of blood.   A standard type of instrument is the Bingham viscometer.   The technique of measurement with this viscometer has been described by Swindells (9)

Clinically, it is advantageous to have a viscometer that can measure very small samples of blood.   The Hess viscometer can measure the viscosity of a single drop of blood.   This apparatus can be obtained from most of the well known apparatus companies.   It consists essentially of two capillaries of equal bore and equal length connected by a T tube with a suction bulb.   Simultaneously, blood is sucked through one capillary and water through the other.   The relative viscosity of the blood as compared to the water is determined from the volume of blood that has flowed through one capillary in comparison with the volume of water that has flowed through the other capillary, the viscosity being inversely proportional to the volume of flow for a given time.

Figure 1 shows the essential tubes of the apparatus.   The two

sets of tubes are filled from the right end up to the zero marks. Then
the same suction is applied to both sets of tubes by means of the bulb
G. When the blood reaches mark 1 (or ½ or 2) in tube M_2, the suc-
tion is cut off, and the position of the water meniscus in tube M_1 is
read. The numbers on tube M_1, which is wider than tube M_2, repre-
sent volumes of water in the same units as do the numbers on tube
M_2. If the blood in tube M_2 is at mark 1, the number at the water
meniscus in tube M_1 is the relative viscosity. Further details of
technique may be found in an article by Bircher (10).
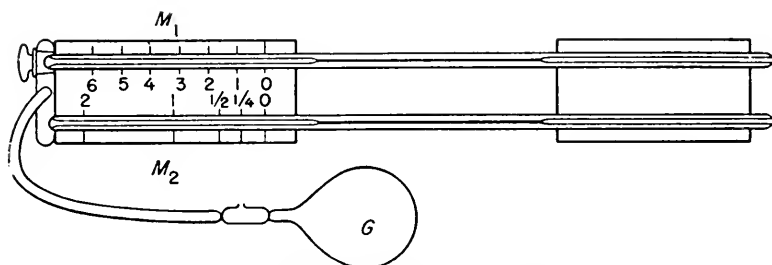
Fig. 1.   The Hess viscometer (after Barr, 3).

For clinical studies, no pressure gage is necessary, but if tests at
different rates of shear are desired, the apparatus can be equipped with
an air reservoir and a gage.

D. VISCOSITY OF PROTOPLASM

Occasionally, protoplasm occurs in rather large masses and it is
possible to obtain quantities of living substance that can then be ex-
amined in the same way that any fluid can be studied. Thus, the
protoplasm of slime molds can be obtained in sizable drops and these
drops can then be forced through a capillary viscometer. This was
done by Pfeiffer (11). Pfeiffer also studied protoplasm squeezed out
of cells of the alga *Chara*, although in this case the protoplasm must
have been mixed with cell sap. In the case of the squid giant nerve
fiber, the protoplasm readily flows out when the fiber is cut. Such
protoplasm can also be forced through a capillary tube, and, in this
way, the viscosity can be studied. Some preliminary studies in this
direction were made by a former student of mine, Mr. L. Nelson, but
nothing has as yet been published. Undoubtedly, when protoplasm
is squeezed out of the cells in which it normally occurs, it undergoes

changes in its physical state, so that any data obtained from extruded
protoplasm must be regarded with caution, and no certain conclusions
can be drawn from such data as to the viscosity of the protoplasm
before it was squeezed out of the cells that contained it.   Some types
of plant protoplasm are in a continual state of flow within their cell
walls, and it might be thought that, from a study of the rate of such
movement, information could be gained as to the relative viscosity
of the protoplasm under a variety of conditions.   Unfortunately,
however, almost nothing is known as to the nature or magnitude of
the forces that govern such protoplasmic flow, so that a change in the
rate of flow following exposure to a given agent or condition might be
due either to a change in the propelling force or to a change in the vis-
cosity (or both).

   At the present time, there are only two standard types of method
commonly used for measuring protoplasmic viscosity.   In one type
of method, granules or inclusions are moved through the protoplasm
by gravity or by centrifugal force.   In the other type of method, the
speed of Brownian movement is used as a measure of viscosity.   Heil-
bronn (12) has employed a magnetic method.   He inserted small iron
wires into the protoplasm of slime molds and then used an electro-
magnet to rotate them.   This method is only applicable to slime mold
protoplasm (or perhaps also the protoplasm of the giant nerve fiber
of the squid).

1. Gravity and Centrifuge Methods

   The two most widely used methods for studying the viscosity of
fluids are the capillary tube method and the falling sphere method.
As has been noted, the tube method is not appropriate for living
uninjured protoplasm.   On the other hand, the falling sphere method
can be applied to many types of living cells.

   The falling sphere method depends on the application of Stokes'
law to the movement of a small sphere through a fluid.   Derived
mathematically many years ago, Stokes' law has been widely used
in various fields of physics and physical chemistry.   In its original
form Stokes' law is:

$$W = 6\pi\eta a v$$

in which W is the force pushing the sphere through the fluid, η is the
viscosity of the fluid, a the radius of the sphere and v its velocity.   If

the sphere is falling under the influence of gravity, the force acting
on it is its apparent weight, that is to say, its volume times the dif-
ference between the specific gravity of the sphere, σ, and the fluid, ρ,
times the gravity constant, g.   Thus:

$$W = \tfrac{4}{3}\pi a^3(\sigma - \rho)g = 6\pi\eta a v$$

so that:

$$\eta = 2g(\sigma - \rho)a^2/9v$$

and:

$$V = 2g(\sigma - \rho)a^2/9\eta$$

Thus, if we have a sphere falling through a fluid and we know its
specific gravity, its radius, its speed, and the specific gravity of the
fluid through which the sphere falls, we can know the viscosity.

In the original derivation of the law that bears his name, Stokes
assumed that certain mathematical terms, the so-called semiquad-
ratic terms, could be neglected.   Actually, they can properly be neg-
lected only if the movement of the sphere is slow and its size small.
Thus Rayleigh in 1893 pointed out that the assumption is warranted
only if $v a \rho/\eta$ is negligible compared to unity.   This condition is read-
ily met in studies in protoplasmic viscosity, for in protoplasm the
spheres that move are tiny, of the order of magnitude of 10^{-4} cm.,
and their rate of movement very slow, typically 10^{-4} cm. per second.
Modern authors agree that the original form of Stokes' law holds if the
Reynolds number $(2v a \rho/\eta)$ is small.   Barr (3) states that, so long as
the Reynolds number is negligibly small compared with unity, the
original Stokes law holds; if the Reynolds number is less than 0.05,
Stokes' law is accurate to about 1%.   Schiller (13) also states that
Stokes' law holds if the Reynolds' number is much less than one.   At
higher speeds, when the Reynolds' number is larger, a more exact
formulation has been derived by Oseen.   In the studies of protoplasm,
Oseen's law is of no advantage, and the original form of Stokes' law
can properly be used.

The theoretical derivation of Stokes' law involves a number of
assumptions.   These are that the motion of the sphere is slow and
that the motion is steady and free from acceleration, that there is no
slip between the sphere and the fluid, that the sphere is rigid, and
that the fluid is homogeneous and extends infinitely in all directions.
As Arnold (14) showed, these assumptions do not affect the validity
of the law for a small particle dropping through a viscous fluid (cf.
also the discussion given by Barr, 3).   For narrow tubes, a correction

is necessary, but such correction becomes unimportant if the diameter of the tube is more than ten times that of the sphere (*4*).

The application of Stokes' law to protoplasm is discussed by Heilbrunn (*1,15*). In the study of protoplasmic viscosity, the spherical particles that move through the protoplasm are sufficiently small and their movement is slow. Nor is one ordinarily troubled by the need for corrections due to the walls of the cell. On the other hand, there is one serious difficulty. This is the fact that Stokes' law governs the movement of a single sphere. The presence of more than one sphere introduces complications. In protoplasm, there are numerous particles that move under the influence of gravity or centrifugal force, and the velocity of any single particle tends to be slowed by the presence of other particles in its neighborhood. Various attempts have been made to apply Stokes' law to the movement of large numbers of spheres, and corrections to the law have been proposed by Cunningham, Smoluchowski, and others. Of these corrections, the only one that has been used in the study of protoplasm is that of Cunningham (*16*). When the Cunningham factor is introduced into Stokes' law, this becomes:

$$v = \frac{2g(\sigma - \rho)^2}{9q\eta}$$

in which:

$$q = \frac{4b(b^5 - a^5)}{(b - a)^2(4b^4 - b^3a - 6b^2a^2 - ba^3 - 4a^4)}$$

a is the radius of the sphere; b is half the distance between two adjacent spheres. But, in making the computation, Cunningham finds it is better to substitute for b a quantity b', which is equal to $b\sqrt{3/2}$. As a matter of fact, in most determinations of protoplasmic viscosity, the biologist or biophysicist is interested in the changes of viscosity under different conditions rather than in the absolute viscosity. For measurements of relative viscosity, no attention need be paid to the Cunningham correction.

Gravity Method. Because of the fact that in cells generally the particles do not fall under the influence of gravity, the gravity method has been used only rarely. The technique of the method is extremely simple. All that is necessary is to watch the fall of the moving particles with a microscope. The microscope is set up so

that the stage is vertical (and the tube horizontal).   In the original use of the gravity method, the German botanist Heilbronn (*17*) observed the rate of fall of starch granules through the protoplasm of root cells in a bean plant.   He then compared the rate of fall of the same particles in distilled water.   He regarded the ratio of the speed of fall in the two media as a relative measure of the viscosity of the two media.   Actually Stokes' law should have been employed.   The fall of a starch grain through protoplasm is slower than it is through water, not only because of the greater viscosity of the protoplasm, but also because of its greater specific gravity.   Moreover, because starch grains are relatively large in comparison with the diameter of the tube through which they drop, a correction must be made for the effect of the container wall.   In addition, another correction is necessary because of the fact that many starch grains are falling at the same time.

There are two other instances in which the gravity method has been used.   In the immature eggs of many animals, there is a large nucleus, the germinal vesicle, and this large nucleus contains a spherical nucleolus.   In the egg of the lobster, it was found many years ago that the nucleolus falls under the influence of gravity. Similarly, the nucleolus can be seen to fall in other types of marine eggs.   Gray (*18*) measured the speed of fall in the immature egg of the sea urchin.   From this data, on the basis of reasonable assumptions as to the specific gravity of the nucleolus and the nuclear fluid, Heilbrunn (*1*) calculated the viscosity of the nuclear fluid.   Harris (*19*) has made accurate photographic measurements of the rate of fall of the nucleolus, and has recalculated the viscosity on the basis of these measurements.   Because of the fact that the nucleolus is a single sphere, relatively small in size in comparison with its containing nucleus, Stokes' law can very properly be used for a determination of the viscosity of the nuclear fluid.   It is necessary, however, to make assumptions as to the specific gravity both of the nucleolus and of the nuclear fluid.   For relative measurements of the viscosity of the nuclear fluid under different conditions, no such assumptions have to be made.   All that is necessary to assume is that the specific gravities do not change markedly during the course of the experiment. Also, within the past few months, Rieser (*19a*) has been able to use the gravity method for isolated muscle fibers.   Small oil drops were injected into these fibers.   In most cases the injection itself causes injury and the oil drops do not move through the protoplasm under

the influence of gravity.   But sometimes, when injury is at a mini-
mum, the injected oil drops rise through the protoplasm, and in these
instances, by measuring the speed of movement, viscosity determina-
tions can be made.

Centrifuge Method.   This method has been much more widely
used than the gravity method.   The principle underlying the
method is the same as for the gravity method, except that the move-
ment of particles through protoplasm is under the influence of cen-
trifugal force rather than gravity.   In using Stokes' law, therefore,
we must substitute for g the term cg, in which c is the centrifugal
force in terms of gravity.   The formula then becomes:

$$r = \frac{2cg(\sigma - \rho)a^2}{9\eta}$$

For a given centrifuge, c can be calculated if we know the number of
turns per second (n) and the radius of the circle described by the ends
of the centrifuge tubes (r).   Then:

$$c = 4\pi^2 n^2 r/g$$

One advantage of the centrifuge method is that it makes possible
measurements of viscosity at different rates of shear, for it is only
necessary to vary the centrifugal force in order to obtain a wide range
of forces.   This has both a theoretical and a practical advantage.
Often, it may be desirable to know whether or not the protoplasmic
viscosity varies with rate of shear.   Practically, one chooses a cen-
trifuge that gives a convenient speed of movement of the protoplasmic
granules.

Generally, the best type of centrifuge to use is a hand centrifuge.
Electric centrifuges do not start or stop promptly, and with them it is
scarcely possible to apply a known amount of centrifugal force for
short time intervals.   Years ago, various types of hand centrifuges
were on the market.   Now, it is difficult to obtain hand centrifuges
that turn rapidly enough.   Such centrifuges may be obtained from
the J. H. Emerson Co., Cambridge, Mass.   Any type of hand cen-
trifuge used should be equipped with some sort of a protective device
in case of tube breakage.   The head of the centrifuge should be ar-
ranged so as to hold small glass tubes.   Ordinarily tubes of about 4–5
mm. outside diameter are used.   It is important to have the tube
free from glass dust.   The ends of the tubes are sealed in a flame some
time before they are ready for use.   If soft glass is used, this can be

done very rapidly. A large supply of tubes is needed and care should be taken so that the tubes of any given pair are of approximately equal weight.

Some years ago, Harvey (20) described a microscope centrifuge, and, for a time, this apparatus was manufactured commercially. With Harvey's centrifuge, it is to some extent possible to watch granules moving through a cell under the influence of centrifugal force. Unfortunately, however, the cells being observed rarely stay in the field, so that only occasionally is one able to make measurements of the speed of movement of the granules.

With an ordinary centrifuge, it is necessary to make series of tests in order to determine the speed of the moving granules. Thus, for example, in a recent study of the viscosity of the protoplasm of the egg of the worm *Chaetopterus* (20a) the centrifuge was turned for periods of 4, 5, 7, 9, 10, 11, 12, 13, 14, and 15 seconds. At the expiration of each test, the eggs were removed and examined under a microscope. That number of seconds which caused movement of the granules for a given distance could then be recorded.

For some materials, when relatively high centrifugal speeds are used, the eggs may be thrown too violently against the end of the tube, and injury can result. To prevent such injury, the ends of the tubes may be filled with a small drop of a liquid of specific gravity somewhat higher than that of the cells being centrifuged. For example, when marine eggs are studied, it is convenient to place in the ends of the tubes a solution of sucrose approximately isotonic with sea water. A 0.95 molal solution is frequently used for this purpose. For nonmarine material, such sugar solutions are too hypertonic, and solutions of gum arabic can be used to cushion the cells. It is a remarkable fact that cells may be exposed to strong centrifugal force without injury. Some marine egg cells after fertilization will continue to develop while being vigorously centrifuged. Cells not surrounded by a stiff membrane when centrifuged vigorously for long times may separate into halves as a result of the opposing pulls of light and heavy granules moving in opposite directions. But, until such breaking occurs, the cells behave normally.

With the centrifuge method it is a simple matter to obtain relative viscosity measurements. The protoplasmic viscosity under one set of conditions is compared to the viscosity under another set of conditions. Measurements of absolute viscosity are much more difficult. For such measurements it is necessary to know the specific gravity of

the moving granules as well as the specific gravity of the fluid through which they move. In order to determine the specific gravity of granules, cells are crushed and the granules are obtained in free suspension. The specific gravity of the granules is then determined by centrifuging them in various strengths of sucrose solution of known specific gravity. In a similar way, one can determine the specific gravity of the cells as a whole. Then, if one knows the fraction of the cell occupied by granules, and this can readily be determined, one can calculate by simple algebra the specific gravity of the fluid through which the granules move. As a matter of fact, determination of the specific gravity of granules is not too exact, for it involves the assumption that, in the sugar solution, the granules do not change their specific gravity. The granules probably do change their specific gravity while exposed to the hypertonic sugar solution. However, the error involved is not very great, and presumably it is not so great as the error involved in determining the radius of the granules. For such small objects as cell granules, it is not possible to arrive at an exact value for their diameter. Unfortunately, the value that is used is squared in the Stokes' formula. Another difficulty that lies in the way of exact determinations of absolute viscosity is the fact that the Cunningham correction, which must be applied (see above), is not a very accurate correction.

All in all, any measurements of the absolute viscosity of protoplasm that have been obtained are not very accurate. Nevertheless, for the two cases in which measurements of protoplasmic viscosity have been made by both centrifuge and Brownian movement methods, reasonably good agreement has been obtained. These measurements of absolute protoplasmic viscosity have shown that the old notion that protoplasm is a very viscous fluid is wrong, at least for some types of protoplasm. Older authors had frequently emphasized the high viscosity of protoplasm, and some indeed had assumed an arbitrary value of 1000 centipoises. Now we know that some types of protoplasm have a viscosity of approximately 3 to 5 centipoises.

In some types of cells, various types of granules move under the influence of centrifugal force. This is a help to the student of protoplasmic viscosity. In the egg of the sea urchin *Arbacia* (see Fig. 2), there are several types of granules that move centrifugally. Chief among these are the small colorless granules. Some larger red granules, several times as large as the colorless ones, also move centrifugally. Because of their larger size, the red granules move more

rapidly through the cell, pushing their way through both the proto-
plasmic granules and the smaller granules in their path.   The speed
of these larger granules can be used for a determination of the viscos-
ity of the protoplasmic suspension (fluid plus smaller granules).   The
value obtained for this viscosity is several times greater than the
viscosity value for the fluid through which all the granules move.
This is what one would expect from our knowledge of the viscosity of
suspensions.
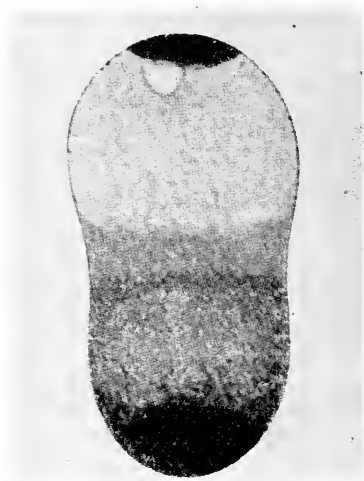
Fig. 2.   A centrifuged sea urchin egg (15).   The egg
is almost ready to break into two halves.

The fact that there may be several types of granules moving
through a cell is a help also in determinations of relative viscosity.
During the mitotic division of a cell, the viscosity of the protoplasm
changes markedly; it becomes more viscous before the appearance
of the mitotic spindle and then becomes rapidly less viscous.   It might
be argued that these changes as observed by the centrifuge method
are not real and that the greater or lesser speed of the granules moving
through the cell is due to changes in the specific gravity of the gran-
ules.   For, according to Stokes' law, if the granules that move centrif-
ugally were to become heavier, they would move more rapidly.   The
fact that the movement of all types of granules, heavy and light alike,
agree in showing the same types of viscosity change is strong evi-

dence that the changes are real, for it would be hard to assume that, whenever the heavy granules became heavier, the light granules became lighter to the same degree.

Although most of the information we now have about the viscosity of protoplasm has been obtained by the centrifuge method, the unfortunate fact is that relatively few types of cells can be studied by this method. The egg cells of various invertebrates, echinoderms, worms, mollusks, etc., frequently provide excellent material; but not always. Thus, for example, although the protoplasm of the egg of the sea urchin *Arbacia* stratifies beautifully and various other types of sea urchin eggs behave in similar fashion, the eggs of the European sea urchin *Paracentrotus* are apparently so full of granules that these granules cannot be moved into one part of the cell. When this egg is centrifuged for long times, only a small region of the cell becomes free of granules. It should be remembered that spherical particles of uniform size cannot fill more than three-quarters of a given space; actually 74.05% is the maximum. Presumably some cells have concentrations of granules that approach this percentage. This seems to be true for the *Paramecium* cell. Paramecia may be centrifuged more vigorously for long times, but the granules in the protoplasm do not move. However, viscosity measurements can be made for *Paramecium* protoplasm by studying the centrifugal movement of food vacuoles. Various types of materials can be fed to the *paramecium*, for example, Chinese ink particles, carmine particles, or starch grains.

Amoeba is a very useful organism for protoplasmic viscosity studies. It is the only type of cell in which studies can be made of the viscosity of both the outer cortex of the cell and the interior protoplasm. For such studies, two species of amoebae are used. In *Amoeba dubia*, there is a very thin cortical layer, and when this amoeba is centrifuged, one can study the viscosity of the main mass of the protoplasm. On the other hand, the related species, *A. proteus*, has a relatively thick cortex. When this amoeba is centrifuged, it is easy to study the viscosity of the stiff cortex, but because this cortex is thick, it is not possible to study the interior. By making studies of both types of amoebae, it is possible to obtain information as to the viscosity of both cortex and interior. In the unfertilized sea urchin egg, the cortex is very thin; indeed it is only a single layer of granules thick. The viscosity of this cortex is so high that it has not been possible to make viscosity determinations for it. Sometimes

the thickness of this viscous cortex increases at the expense of the protoplasm in the interior. Some agents, indeed, cause a progressive thickening of the cortex, until most of the protoplasm becomes cortex.

Although egg cells and some protozoa are suitable for centrifuge studies, many other types of cells are not. Attempts have been made to study the viscosity of protoplasm of frog muscle cells by the centrifuge method—without success. Epithelial cells are small and thus difficult to study. There is no published work on the centrifugal force needed to move granules through various types of nerve cells. Many plant cells might provide suitable material. The chloroplasts can be moved through the protoplasm of the cells of the leaves of the water plant *Elodea*. In some algae, various structures can be seen to move through the cell under the influence of centrifugal force. Thus, Northen (*21*) observed the movement of chromatophores in *Spirogyra*. Presumably, when these structures are moved through the cell, their movement is due to a tearing loose from attachments. Stokes' law can scarcely be used for such a movement and, although the information obtained by Northen is of interest in the interpretation of physical changes in protoplasm, it can hardly give direct or certain information as to viscosity change in the main mass of the protoplasm.

2. Brownian Movement Method

Brownian movement of living material has been known for a long time. As a matter of fact Brown's discovery of the motion was really a discovery of the fact that movement could occur in nonliving as well as living material. Soon after the middle of the nineteenth century, students of protoplasm made observations on the Brownian movement of particles within the cytoplasm and the nucleus, and some of these workers recognized clearly that the rate of this movement was an index of the fluidity of the protoplasm. But, in general, the earlier observations were not very exact. Various conditions were seen to affect the rapidity of the movement; and indeed sometimes the movement seemed to cease entirely. Such cessation of movement could be taken to indicate a marked increase in the viscosity of the protoplasm.

Fortunately, by the application of Einstein's equation for Brownian movement, it has been possible to obtain more exact estimates of

the viscosity. Thus Heilbrunn (1928) studied the rate of return of granules centrifuged to one-half of a sea urchin egg cell. These granules return to their original position by virtue of their Brownian movement. According to the Einstein equation for Brownian movement:

$$D_x{}^2 = (R/N)(Tt/3\pi\eta a)$$

in which D_x is the distance traveled by the granule in any given plane, R is the gas constant in c.g.s. units, T the absolute temperature, t the time, η the viscosity, a the radius and N the Avogadro number.

In order to use the Einstein equation it is not necessary to know the specific gravity either of the granules or the fluid through which they move.

The Einstein equation was also used by Baas-Becking, Sande Bakhuyzen, and Hotelling (*22*) in their study of Brownian movement of particles in the protoplasm just under the cell wall in the alga *Spirogyra*. Using a magnification of 2100×, they made numerous records of the movement of individual particles. Such records are very hard to obtain. "The position of a certain particle was recorded on coordinate paper every 15 seconds. At this high magnification the procedure is strenuous; 180 seconds being the physical limit of observation. Moreover, the particle often disappears behind other structures." It might be thought that it would be simpler and more accurate to use motion picture records, but Baas-Becking, Sande Bakhuyzen, and Hotelling state that such records would necessitate the use of very thin preparations and that the Einstein equation does not hold for such thin films.

As a matter of fact, the movements observed by Baas-Becking, Sande Bakhuyzen, and Hotelling were in thin films. The protoplasm under the cell wall in the *Spirogyra* cell is a very thin layer bounded by two concentric cylindrical surfaces. The particles observed by Baas-Becking, Sande Bakhuyzen, and Hotelling were only 0.4 μ in diameter and their movement was rapid. Obviously, they must have collided frequently with the limiting surfaces of the thin layer of protoplasm in which they were enclosed. Such collisions would tend to restrict the amplitude of the movements and they would tend to make the speed of movement variable and frequently less than it would have been had no limiting surface been present. Accordingly, it is not surprising that highly variable values for the absolute viscosity were found by Baas-Becking, Sande Bakhuyzen, and

Hotelling.   Probably, their minimal values for protoplasmic viscosity
represent approximations to true values.   Also their relative values
for viscosity at different temperatures are significant.

Doubtless the best measurements of viscosity by the Brownian
movement method are those of Pekarek (*23*).   He follows a procedure
used first by Fürth (*24*).   Instead of measuring the exact path of
movement of a given particle, Pekarek determines the time it takes
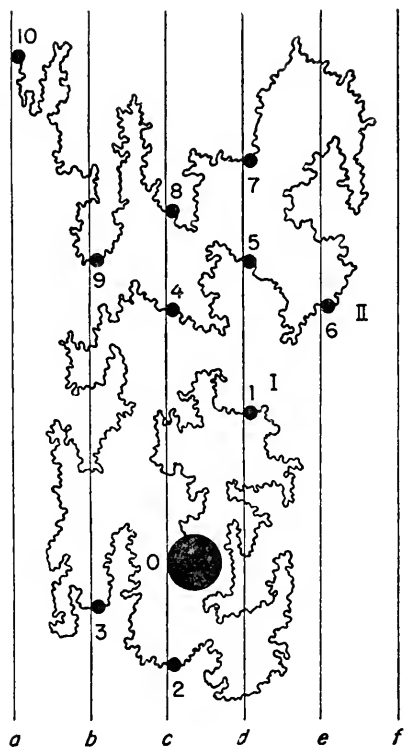
Fig. 3.   Path of a particle in Brown-
ian movement (after Pekarek, *23*).
This is diagrammatic, not an actual
record.   Numbers 1–10 indicate "pas-
sages" according to Pekarek's method
of counting.

for a particle to cover a given distance to the right or left of its original
position.   A microscope ocular with a finely divided series of parallel
lines is used.   Each time the particle in Brownian movement covers
the distance between two sets of lines, this is considered a "passage."
Pekarek determines the number of passages for a given time interval.
In counting passages, a particle must pass over a complete space be-
tween two lines.   This is illustrated in Figure 3, which shows the

movement of a particle during ten passages. If the number of passages for a given time interval is known, the viscosity can be computed from the following formula, due originally to Fürth:

$$\eta = \frac{R}{N} \frac{Tt}{3\pi a l^2 n}$$

In which η is the viscosity, R the gas constant, N the Avogadro number, T the absolute temperature, t the time interval for the total number of passages, n, and l the distance traveled for a given passage.

Pekarek tested the method for distilled water and found excellent agreement with the known value for the viscosity. In using the method, certain conditions must be fulfilled. The cells examined must be intact and healthy. In the protoplasm studied, there should be no directed movement, such as occurs in protoplasmic streaming or in the so-called "Glitschbewegung" found in the cells of certain algae. The Brownian movement should be random; this necessitates making a sufficient number of observations. Inasmuch as the Fürth formula depends on the Einstein formula for Brownian movement, and this in turn depends on Stokes' law, the limitations of Stokes' law should be noted. The fluid through which the granules move should be homogeneous, that is to say, the diameter of the particles should be large in comparison with the mean free path of the fluid molecules. This condition is easily satisfied. Then, too, the particles should be spherical. They should not change size during the course of a determination. The presence of a limiting surface or wall in the vicinity of the moving particles introduces complications; hence it is wise to study particles in relatively large masses of protoplasm. Finally, the number of particles should not be too great, for if the particles are crowded their movement is retarded.

In view of these various difficulties, it is not a very simple matter to find material suitable for study. Pekarek determined the viscosity of the cell sap in the vacuoles of various plants. Such measurement was relatively easy by his method. In his studies of actual protoplasm, Pekarek determined the viscosity of the protoplasm of an amoeba that did not contain too great a concentration of granules and was not actively moving. He also studied the protoplasm of the rhizoids of the alga *Chara*. For *Amoeba* protoplasm, he obtained a value of 5 centipoises, which is exactly the same as that previously found by Heilbrunn (25) using the centrifuge method. For *Chara* protoplasm, Pekarek also found the same value.

E. RESULTS OBTAINED FROM VISCOSITY STUDY

Numerous authors have studied the viscosity of the blood of various mammals. Essentially the same values have been obtained for the past forty or fifty years. Ordinarily, in recent studies at least, the Hess viscometer is used. With this apparatus, as noted previously, measurements can be made at various pressures, but usually only a single pressure is used, and the viscosity of the blood at this pressure is compared with the viscosity of water.

Blood plasma and blood serum obey Poiseuille's law, but whole blood, to some extent, shows an anomalous viscosity. The variation from Poiseuille's law only becomes significant for very low shearing forces; this variation tends to become more evident if the capillary tube through which the blood flows has a small diameter (see Hess, 26). In making comparisons of the rheological properties of various types of bloods, Bingham and Roepke (27) prefer to use the fluidity value instead of the viscosity value. The fluidity is the inverse of the viscosity, and its unit, the rhe, is $1/\eta$. Recent data on the viscosity of the blood of various mammals and birds are given by Rhiel (28) and Moll (29).

In clinical diagnosis the viscosity of the blood is sometimes measured, although such measurements are not thought to have as much diagnostic value as the results obtained from the other types of measurements. Usually the normal viscosity of the blood (in terms of the viscosity of water) is taken as 4.5. In cases of anemia and nephritis, the viscosity is lower, and it is usually lower also in leukemia and malaria. In diabetes mellitus and in jaundice, the viscosity is low, and this is usually true also in pneumonia. In general, although not always, when the volume of red blood cells in a given volume of blood is high, the viscosity of the blood is greater (cf. 30). Under conditions that induce excessive sweating, most of the loss of water is from the blood, and such water loss is accompanied by increased viscosity. In traumatic shock also, there is typically a loss of water from the blood. Such changes in the total blood volume are usually studied by measuring the concentration of a dye injected into the blood stream, by hematocrit studies of the volume of blood corpuscles in a given volume of blood, by blood cell counts, by refractometer measurements or by hemoglobin determinations. Physiologists and clinicians are more accustomed to such techniques than they are to viscosity measurements, so that ordinarily, in physiological experimen-

tation and in clinical practice, no measurements of blood viscosity are made.

It is obvious that any increase in blood viscosity would impose an added strain on the heart, and this may be an important factor in some types of distress or disease.

Studies of protoplasmic viscosity have given far more interesting results than studies on blood viscosity. Earlier biologists and physiologists were forced to guess as to the changes in viscosity and colloidal state. Muscle contraction was thought by some to involve a sharp increase in protoplasmic viscosity—a gelation—whereas others decided on the basis of reasoning alone that muscle protoplasm changed from a viscous and elastic gel to a fluid sol when it was made to contract. Similarly, in other processes, such as cell division, changes in viscosity were assumed rather than measured. The action of drugs was explained in terms of viscosity change. Thus Claude Bernard believed that ether and chloroform produced their effect by causing a semicoagulation of the protoplasm.

The introduction of correct methods of protoplasmic viscosity has replaced the earlier speculations with definite facts. A summary of this factual information is given by Heilbrunn (1,15). As already noted, some types of protoplasm have a low viscosity. The values of the absolute viscosity of such protoplasm are approximately 3–5 centipoises. This is the viscosity of the nongranular protoplasm The viscosity of the entire protoplasm, granules and all, depends on the concentration of the granular suspension. In the sea urchin egg, the entire protoplasm has a viscosity several times as great as that of the hyaline protoplasm. But, where the concentration of granules is very high, the viscosity of the entire protoplasm may be very great. Thus in *Paramecium* protoplasm, the protoplasmic viscosity may be several thousands of centipoises.

The interior protoplasm of a cell may have a low viscosity, whereas the cortical region of the cell may have a very high viscosity, even approaching infinity. This outer cortex may be very thin, as in some marine eggs, or it can be much thicker, as in the common *Amoeba*. Perhaps in some cells essentially the entire protoplasm has the properties of a cortex.

In *Amoeba*, the high viscosity of the cortical protoplasm seems to depend on the presence of calcium there. If this calcium is in part removed by sodium or ammonium oxalate, the viscosity of the cortex decreases sharply, and a similar decrease occurs if the calcium of the cortex is replaced by potassium as a result of ion exchange.

When an amoeba is exposed to stimulating agents, such as an electric shock, ultraviolet radiation, or mechanical agitation, the viscosity of the cortex drops sharply, due apparently to a release of calcium. The released calcium apparently enters into the interior of the cell. There it causes first a drop in protoplasmic viscosity, followed by a sharp increase in viscosity.

Such a sharp viscosity increase seems to occur in all cells when they are exposed to stimulating agents. Even nerve protoplasm appears to become much more viscous when stimulated electrically (*cf. 31*). The effect is thought to be due to a clotting reaction similar to blood clotting, and dependent primarily on the entrance of free calcium ion into the cell interior. It is thought that, when a muscle is stimulated to contract, when a nerve is excited, and when an egg cell is induced to divide, in all cases a gelation or protoplasmic clotting occurs and that this clotting is similar in its fundamental aspects to blood clotting.

Fat solvent anesthetics tend to prevent the clotting reaction of protoplasm. Apparently, their action is to prevent binding of calcium either in the cortex or the interior of the cell. They cause a liquefaction of the cortex and they tend to prevent gelation of the interior.

In addition to the fat solvent anesthetics, various drugs and other agents have a marked effect on protoplasmic viscosity. At the present time, we know something as to the effect of hypotonic and hypertonic solutions, acids and alkalies, metallic poisons, heat, cold, radiation, pressure, etc. On the basis of the information already obtained, it is possible to draw various conclusions as to the behavior of the protoplasmic colloid. Indeed our knowledge of the colloid chemistry of protoplasm is very largely based on quantitative studies of protoplasmic viscosity.

F. POSSIBILITIES FOR THE FUTURE

The most interesting viscosity change in blood is undoubtedly the change that occurs when blood clots. This change is not studied by the ordinary methods of viscometry, but is followed by observing the length of time it takes for shed blood to transform from a fluid sol to a gel with some of the properties of a solid. Clotting times of blood can easily be measured in large tubes, drops, or capillary tubes.

Unfortunately, the student of blood has devised no quantitative methods for the study of blood clotting within the intact circulation.

The flowing blood normally contains thromboplastic substances and substances like thrombin, which promote clotting, and substances like heparin and fibrinolysin, which tend to prevent or reverse clotting. Within the intact organism, there is constant danger of blood clotting within the vessels as thrombin and thromboplastic substance increase in amount. It is probable that small clots frequently form within the blood stream and are reversed or liquefied before they cause serious damage. Ordinary methods of blood viscometry do not reveal the existence of incipient or partial clots in the blood. For, in order to obtain samples for viscosity tests, the blood is either defibrinated or it is treated with anticoagulants. Such treatment would mask the existence of clots within the blood.

When tissues of the body are injured, beyond much doubt they throw into the circulation injury substances that have a thromboplastic action and would therefore tend to induce clotting. This may be, and probably is, a primary factor in the cause of traumatic shock. It would be interesting, therefore, to know how the viscosity of the flowing blood changes following tissue injury. The studies of Knisely (32) clearly indicate that, following trauma, there is a marked increase in blood viscosity, presumably because of incipient or partial clotting. Knisely describes what he calls a sludging of blood as a result of tissue injury. Knisely's studies are primarily morphological and are the result of microscopic study. They emphasize the importance of attempting to make viscosity measurements of blood within the organism or at least without the addition of anticoagulants.

Although for the past thirty years the interest in protoplasmic viscosity has been steadily maintained, there has been no great enthusiasm for the field. For each paper on the viscosity of protoplasm that has been published, there have been at least dozens on cellular respiration. And this in spite of the fact that the correlations between protoplasmic activity and viscosity appear to be much more satisfactory than any correlations between such activity and respiration. This has certainly been shown true in the study of such phenomena as cell division, anesthesia, and drug action. The reason for the lack of literature in the field of protoplasmic viscosity is easy to understand. It is hard, at least in animal tissues, to find cells suitable for viscosity studies. Until recently no one has been able properly to measure the viscosity of muscle protoplasm. Fortunately, as noted earlier, Rieser has developed an accurate method for making such determinations (19a). In the future this will doubtless yield

important information. As yet, there have been no accurate studies of nerve cells, but conceivably the centrifuge method can be used on the giant nerve cells of annelids. In the past the only animal cells that have given results that have been useful were marine egg cells and protozoan cells. The egg cells can be studied only at marine laboratories and then only at certain seasons of the year. Protozoan cells are often difficult to handle. If a careful search is made, it should be possible to discover new types of material for measurements of protoplasmic viscosity. If this material can be found, the future for the field would be very bright indeed. The colloidal behavior of protoplasm is best studied by viscosity measurements. If we knew more about the colloid chemistry of protoplasm, we would be in a position to provide basic information useful not only for the biologist and physiologist, but also for the pathologist, the pharmacologist, and even for the clinician. For surely knowledge as to the nature and behavior of the protoplasmic colloid is of primary importance for all sciences concerned with the behavior of living organisms.

References

1. Heilbrunn, L. V., *The Colloid Chemistry of Protoplasm*. Borntraeger, Berlin, 1928.
2. Heilbrunn, L. V., and F. J. Wiercinski, *J. Cellular Comp. Physiol.*, **29**, 15 (1947).
3. Barr, G., *A Monograph of Viscometry*. Oxford Univ. Press, London, 1931.
4. Bingham, E., *Fluidity and Plasticity*. McGraw-Hill, New York, 1922.
5. Hatschek, E., *The Viscosity of Liquids*. Bell, London, 1928.
6. Scott Blair, G. W., *An Introduction to Industrial Rheology*, Blakiston, Philadelphia, 1938; *A Survey of General and Applied Rheology*, Pitman, London, 1943.
7. Reiner, M., *Ten Lectures on Theoretical Rheology*. Mass, Jerusalem. 1943.
8. Tobolsky, A., R. E. Powell, and H. Eyring, "Elastic-Viscous Properties of Matter," in *The Chemistry of Large Molecules*, R. E. Burk and Oliver Grummitt, eds. Interscience, New York, 1943, pp. 125–190.
9. Swindells, J. W., *J. Colloid Sci.*, **2**, 177 (1947).
10. Bircher, M. E., *J. Lab. Clin. Med.*, **7**, 134 (1921).
11. Pfeiffer, H. H., *Protoplasma*, **33**, 311 (1939); **34**, 347 (1940).
12. Heilbronn, A., *Jahrb. wiss. Botan.*, **61**, 284 (1922).
13. Schiller, L., "Fallversuche mit Kugeln und Scheiben," *Handbuch der Experimentalphysik*, W. Wien and F. Harms, eds. Vol. 2, Part 2, pp. 337–387.

14. Arnold, H. D., *Phil. Mag.* **12,** 755 (1911).
15. Heilbrunn, L. V., *An Outline of General Physiology,* 2nd ed. Saunders, Philadelphia, 1943.
16. Cunningham, E., *Proc. Roy. Soc. London,* **A83,** 357 (1910).
17. Heilbronn, A., *Ber. deutsch. bot. Ges.,* **30,** 142 (1914).
18. Gray, J., *Brit. J. Exptl. Biol.,* **5,** 102 (1927).
19. Harris, J. E., *J. Exptl. Biol.,* **16,** 258 (1939).
19a. Rieser, P., *Federation Proc.,* **8,** 131 (1949).
20. Harvey, E. N., *J. Franklin Inst.,* **214,** 1 (1932).
20a. Heilbrunn, L. V., and W. L.Wilson, *Biol. Bull.* **95,** 57 (1948).
21. Northen, H. T., *Protoplasma,* **31,** 1 (1938); *Cytologia,* **10,** 105 (1939); *Plant Physiol.,* **15,** 645 (1940).
22. Baas-Becking, L. G. M., H. L. van de Sande Bakhuyzen, and H. Hotelling, *Verhandl. Koninkl. Nederland. Akad. van Wetenschap., Afdeel. Natuurk.,* **25,** No. 5, 1 (1928).
23. Pekarek, J., *Protoplasma,* **10,** 510 (1930); **11,** 19 (1930); **13,** 637 (1931); **17,** 1 (1932); **18,** 1 (1933); **20,** 251 (1933).
24. Fürth, R., *Ann. Physik,* **53,** 177 (1917); *Z. Physik,* **60,** 313 (1930).
25. Heilbrunn, L. V., *Protoplasma,* **8,** 65 (1929).
26. Hess, W. R., *Arch. ges. Physiol.,* **162,** 187 (1915).
27. Bingham, E. C., and R. R. Roepke, *J. Gen. Physiol.,* **28,** 79, 131 (1944).
28. Rhiel, J., *Arch. ges. Physiol.,* **246,** 709 (1943).
29. Moll, K., *Arch. ges. Physiol.,* **247,** 74 (1943).
30. Eckstein, R. W., D. Book, and D. E. Gregg, *Am. J. Physiol.,* **135,** 772 (1942).
31. Flaig, J. V., *J. Neurophysiol.,* **10,** 211 (1947).
32. Knisely, M. H., *Arch. Surg.,* **51,** 220 (1945).
33. McGoury, T. E., and H. Mark, "Determination of Viscosity," in *Physical Methods of Organic Chemistry,* 2nd ed., A. Weissberger, ed. Interscience, New York, 1949, Chap. VIII.

TEMPERATURE DETERMINATIONS

LAWRENCE R. PROUTY AND JAMES D. HARDY, *Cornell University Medical College*

Temperature is an indication of the thermal energy level of a body. Thermal energy will flow of its own accord from a warm object to a cooler one. Therefore, temperature is a property that determines the rate and direction of this heat flow. A large number of thermal states or temperatures can be differentiated and are characterized by definite phenomena. Any function of matter that varies continuously with temperature may be used to indicate temperature. The most useful of these include the physical phenomena of expansion and change of state, and properties such as the change of electrical resistance of a conductor. The Seebeck effect at the junction of two dissimilar metals is of great practical importance in the measurement of temperature.

A. TEMPERATURE SCALES

1. Arbitrary Scales

Various temperature scales have been devised to characterize the infinite number of thermal states between fixed points. The fixed temperatures most widely used as reference points are the melting and boiling points of pure substances at a constant known pressure. In the scales discussed below the steam and ice points of water were used as reference points.

Celsius Scale. The first thermometric scale was devised by Celsius. He assigned the value of 100 to the ice point and zero to the steam point and divided the interval into 100 degrees. "Degrees Celsius" is a term used in continental Europe at the present time and has come to denote what English-speaking countries refer to as Centigrade temperatures.

Centigrade Scale. The Centigrade scale divides the range between the freezing point of water, arbitrarily called zero, and the boiling point, arbitrarily called 100, at standard atmospheric pressure, into 100 equal divisions each of which is one Centigrade degree. This scale can be employed with the decimal system with facility. The Centigrade scale with several fixed and reproducible equilibrium temperatures is the basis of the International Temperature Scale adopted in 1927 by the General Conference of Weights and Measures. The Centigrade scale is used almost exclusively in scientific laboratories in this country at the present time, and the Fahrenheit scale largely in industry.

Fahrenheit Scale. The same temperature range as in the previously described scales is divided in the Fahrenheit scale into 180 equal parts called Fahrenheit degrees. The value of 32 is assigned to the freezing point of water and 212 to the steam point. It is not known why Fahrenheit used 32 divisions below the freezing point as zero, but it may have represented the lowest temperature he could achieve with an ice–salt combination. It has also been suggested that his zero was the lowest temperature he noted during a particularly severe Danzig winter.

Reaumur Scale. The Reaumur scale divides the same fundamental interval from the ice point to the boiling point into eighty equal intervals of one degree each. This scale is not widely used but is occasionally seen in biophysical papers.

Convenient formulas for conversion from one to another of these scales, where C = Centigrade, F = Fahrenheit, R = Reaumur, are:

$$^\circ F = \frac{^\circ C \times 9}{5} + 32 = \frac{^\circ R \times 9}{4} + 32 \tag{1}$$

$$^\circ C = \frac{(^\circ F - 32) \times 5}{9} = \frac{^\circ R \times 5}{4} \tag{2}$$

2. Absolute (Kelvin) Scale

In measuring temperature by expansion of matter, the magnitude of a degree depends upon the nature of the particular substance used. Thus, the Centigrade, Fahrenheit, and Reaumur scales are intensive rather than extensive scales. In these scales there is no fundamental temperature unit that can be successively applied to measure any other temperature unit as can be done in the measurement of a quantity such as length. The magnitude of a degree on one part of any one of these scales may not be compared to the magnitude of a degree on another part of the same scale. In addition, none of these scales will accurately describe the full range of theoretical temperatures. For example, liquid-in-glass thermometers have the limited temperature range from the freezing point of the expanding substance to the softening point of glass. At low temperatures the phenomenon of "superconductance" renders electrical resistance thermometers useless, and at high temperatures alterations in properties of the metal used cause the resistance to change, even at constant temperature, below the melting point of the metal. Gas thermometry, although

extending over a wider range than liquid-in-glass thermometers, suffers from the same limitations, *i.e.*, the change in state of the gas and the softening point of the gas container.

All measurements should ultimately be made in fundamental units such as mass, length or time, or in derived units such as energy levels. The development of the absolute temperature scale makes possible the measurement of temperature in terms of fundamental units that are independent of the thermometric substance. This scale is useful throughout the entire temperature range. It was formerly believed that heat was a material substance that, when added to a body, made it warmer and, when subtracted, made it cooler. Experimental studies of Rumford (1798), Joule (1842), and Rowland (1873) helped establish the concept that heat is a form of energy and that equivalent relationships may be found between units of heat and units of mechanical energy. Temperature is the property that determines the direction of heat flow.

From thermodynamic considerations, Lord Kelvin proposed a scale of temperature that was independent of the expanding or "working" substance. This scale has more meaning from a physical standpoint since the zero point of the scale corresponds to conditions of zero energy of the working substance and represents the lowest theoretical temperature obtainable. Furthermore, it is a scale to which measurements of radiant energy as well as mechanical energy can be referred conveniently. The absolute scale is a purely theoretical one but, since it corresponds closely with the gas thermometer scale, at temperatures usually encountered, the absolute scale has wide practical application. The size of the degree on the absolute scale was chosen equivalent to the degree Centigrade. Thus:

$$°K = °A = °C + 273$$

The absolute scale is particularly important in those investigations in biophysics involving thermodynamics and heat transfer by radiation.

B. LIQUID THERMOMETERS

1. Miscellaneous Modifications

Alcohol Thermometers. Thermal expansion of alcohol is about ten times as great as that of mercury under similar condi-

tions and, therefore, may be used advantageously as the expansile liquid in a sensitive thermometer below 60°C. In this range an alcohol thermometer having the same sensitivity as a mercury thermometer can have a bulb size one-tenth that of a mercury thermometer of the same size bore. Furthermore, an alcohol thermometer has less tendency to cool the medium being tested than a mercury thermometer and may be used at lower temperatures since alcohol does not solidify until it is cooled to −130°C. However, alcohol boils at 78°C. and cannot be used near or above this temperature except under pressure. The stem of the alcohol thermometer must be kept at the same or at a higher temperature than the bulb because of the possibility of condensation on the inner wall.

Mercury-in-Glass Thermometers. Mercury-in-glass thermometers are the most commonly used temperature-measuring instruments in the laboratory. Mercury, when allowed to expand in only one dimension, expands and contracts linearly with temperature changes between its solidification and vaporization points (−39 to 357°C.). In general, mercury-in-glass thermometers are inexpensive, easily calibrated, and can be obtained in many sizes and ranges for the degree of sensitivity required.

Mercury is confined to a reservoir (bulb) attached to a length of glass tubing (stem), which is sealed in the upper end after evacuation. For use at temperatures between 350 and 500°C. the thermometer is made of special glass with a high melting point and is filled with nitrogen under pressure to raise the boiling point of mercury. Either enclosed (Einschuss) or the more popular etched-stem scales are obtainable. Etched-stem types tend to lose the coloring matter filling the etched lines and become difficult to read, but they are less expensive and less bulky. Enclosed-scale types have the temperature divisions marked on a white background placed behind the stem, both being enclosed in a glass envelope. The large, well protected scale of this type is easy to read but may become loose and slip away from the capillary tubing of the thermometer. Enclosed-stem thermometers need not be discarded if the glass envelope becomes broken since the latter may be replaced.

It is most important to distinguish between accuracy and sensitivity in a thermometer. Sensitivity, the ability of the thermometer to respond to temperature changes, depends upon the ratio of the volume of the bulb to the cross-sectional area of the bore of the tube. To increase the sensitivity for small increments of temperature, the

mercury column must be made to travel over a greater distance along
a stem of small bore from a comparatively large mercury bulb. This
requires either a very long stem or an enlargement in the upper end of
the stem into which the mercury can be driven by heat. Accuracy,
the extent to which the thermometer indicates true temperature
values, can be estimated by the correlation between the temperatures
measured by that thermometer and a precise thermometer in abso-
lute units. It is affected by thermal expansion
coefficients of the glass, sensitivity, uniformity of
bore, ease and skill with which the scale can be
read, preciseness of calibration, and many other
factors.

Any thermometer indicates primarily its own
temperature as affected by the environment. A
thermometer that is sensitive to small temperature
increments will, because of the comparatively large
mass of the bulb, have a greater tendency to modify
the temperature of its environment. Hence, in this
regard, a sensitive thermometer may be less ac-
curate than one of smaller mass and it may be
slower to respond to thermal changes.

Beckmann Differential Thermometers. To
obtain a thermometer that is both sensitive and
accurate, Beckmann devised one (Fig. 1) which,
although it does not permit absolute tempera-
tures to be read, does give differences in tempera-
ture with an error not exceeding 0.001 to 0.002°C.
If this type of thermometer were made in the
ordinary manner, the stem would have to be in-
conveniently long or a number of thermometers
for different temperature ranges would be required.
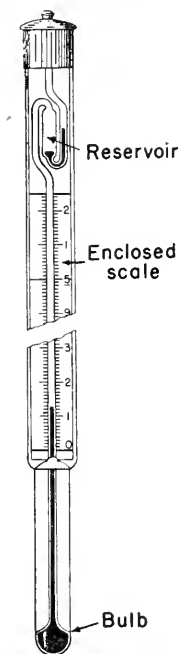
Fig. 1. Beckmann
thermometer.

In the Beckmann thermometer the amount of mercury in the bulb,
and consequently the temperature range of the thermometer, can
be altered by driving some of the mercury into a reservoir connected
to the upper end of the stem. It is necessary to "set" the ther-
mometer before using so that the end of the mercury thread is on
the scale in the desired temperature range. This is done by sus-
pending the thermometer in a beaker of water whose temperature
is regulated with the aid of an ordinary thermometer at a level
two or three degrees higher than the highest temperature to be met

in the experiment. If the top of the mercury column of the Beckmann thermometer does not rise to the scale (*i.e.*, too little mercury in the bulb and too much in the reservoir) the thermometer should be placed in a second bath at a temperature sufficiently high to cause the mercury to pass up and form a small drop at the upper end of the capillary. The thermometer is then inverted and tapped to collect the mercury in the reservoir at the end of the capillary. Next, the thermometer is returned without shaking to the vertical position, and replaced in the first bath. At this time, the mercury in the bulb will contract and draw in some of the mercury from the reservoir. When the thermometer has become fully adjusted to the temperature of this bath the upper end of the thermometer is struck against the palm of the hand in order to break off the excess mercury from the end of the capillary. The thermometer is then tested in the temperature range of the experiment. If the mercury column rises above the scale, too much mercury has been delivered to the bulb; some of it must be driven into the reservoir and a small amount shaken off from the end of the capillary. Many Beckmann thermometers are calibrated so that each of these drops shaken off from the end of the capillary will lower the temperature range of the scale by approximately one or two degrees. These drops of mercury may be left in the glass tubing without returning them either to the bulb or reservoir. Temperature readings are not affected by so doing.

Clinical Thermometers. A clinical thermometer is essentially a maximum-reading, mercury-in-glass thermometer. It is made self-registering by means of a constriction in the bore about 1 cm. above the bulb. This permits mercury to expand upward freely onto the scale as the temperature rises but prevents return of the mercury to the bulb except by vigorous shaking. A clinical thermometer in which the constriction is too small will have the mercury expand upward in a series of jerks and, if too large, will have the mercury retreat into the bulb.

Clinical thermometers usually have an etched stem, and cover a range of 10°F. (96 to 106°F. in intervals of 0.2°) or 6°C. (35 to 41°C. in 0.1° intervals). In order to minimize thermal response time the bulb is made small and consequently the bore must be quite fine. To facilitate reading a white backing is placed behind the capillary (bore) and the clear glass front is molded in the form of a magnifying lens to increase the apparent width of the mercury column.

Maximum–Minimum Thermometers. Maximum and minimum thermometers indicate the highest and lowest temperatures attained during an experimental period. One type of minimum thermometer consists of a small glass rod immersed in alcohol. Due to surface tension the glass rod is drawn along the tube as the alcohol contracts with falling temperature. On expansion of the alcohol with rising temperature, the glass rod remains stationary indicating the lowest temperature reached.

A maximum thermometer is a mercury-in-glass thermometer that may operate on the stem constriction principle (Negretti and Zambra's version) as in clinical thermometers, or, alternatively, employs an index that remains at the highest temperature achieved. This index may be a small piece of iron (Rutherford's version) or a thread of mercury separated from the main column by an air bubble (Phillips' version).

2. Calibration of Mercury-in-Glass Thermometers

The calibration of a mercury-in-glass thermometer is ordinarily accomplished in the laboratory by comparison with a standard thermometer calibrated and certified by the National Bureau of Standards. Thermometers with structural defects that may influence the reproducibility of their readings will be calibrated upon request by the thermometry section of the Bureau of Standards. These thermometers will not, however, be certified. A certificate is supplied with each thermometer calibrated and certified by the Bureau, indicating the correction factors for each stem marking under the conditions of calibration. The certificate also contains information for calculating emergent stem corrections in case it is desired to use a standard thermometer under conditions different from those of the original calibration. Waidner and Mueller (1) studied the relative merits of total and partial stem immersion and concluded that for ordinary laboratory work more accurate measurements can be secured by partial immersion thermometers. Thermometers graduated in divisions smaller than 0.5°C. should usually be standardized as total immersion thermometers. Although the conditions for intercomparison of laboratory and standard thermometers may vary, certain precautions should be taken to insure reproducibility in comparison measurements. Among the most important of these precautions are:

(*1*) Thermal equilibrium between the standard thermometer and the comparison thermometer is essential. This may be provided for by a well stirred water bath.

(*2*) Before testing, the mercury column in both thermometers should be raised to the top of the stem before heating in order to insure continuity of the mercury column. Gentle tapping, with heating and cooling the thermometer bulb, will assist in joining of a separated column.

(*3*) It is generally desirable to make a comparison under conditions of intended use of the laboratory thermometer.

(*4*) Calibrate slowly allowing ample time for establishing thermal equilibrium at several points in the desired temperature range. Calibration from low to high temperature is safer and gentle tapping of the thermometer stems will prevent slight sticking of the mercury column.

(*5*) Parallax errors in reading can be avoided by use of a good thermometer reader.

(*6*) Correction factors are then recorded in tabular form or graphed with temperature on one axis and correction factors on the other. Few thermometers have bores of uniform diameter throughout their length. Thus, the correction factor will vary at the intervals chosen for the calibration. Interpolations are made for corrections between calibrating points.

Beckmann thermometers, in addition to the above calibration factors, must be corrected for a "setting factor" if used in temperature ranges above or below that for which they are calibrated. This is necessary because the quantity of mercury in the bulb is more or less than the amount it contained during calibration.

Clinical thermometers may be calibrated in the same manner as any other mercury-in-glass thermometer but it is usually more important to test the thermometer at least once a year to see if it is a "retreater." Because of an over-sized constriction in the stem, the mercury column retreats into the bulb upon withdrawal of the thermometer from the mouth, rectum, axilla, or skin. Busse (*2*) recommends testing the thermometer by holding it vertically in a beaker of water at a temperature of about 39.4°C. while stirring the water with the bulb. The reading is noted and the thermometer removed and read again while still vertical. If the second reading is more than 0.1°C. lower than the initial reading on several successive tests, the thermometer should be discarded as a "retreater." Clinical thermometers showing a break in the mercury column above the constriction probably have air in the bore and should also be discarded.

The Bureau of Standards accepts as tolerances for clinical thermometers readings that are correct within 0.1°C. at 38°C. and 0.2°

C. at 41°C. This means that the readings of two thermometers with the maximum allowable variation may differ from each other by as much as 0.2°C. at 38°C. and 0.4°C. at 41°C., and shows the importance of using the same thermometer for a patient throughout the course of an illness. Tolerances for other mercurial thermometers may be found in the National Bureau of Standards Circular #8, current edition, on testing of thermometers.

3. Applications of Liquid-in-Glass Thermometers

Within the range of precision that can be obtained with liquid-in-glass thermometers—which is generally sufficient for most biophysical work—this type has the specific advantages of low cost and ease of manipulation beyond that of any other type of thermometer. They are easily calibrated and show little deterioration with age and frequent use. If proper consideration is given to size, sensitivity, and accuracy factors, liquid-in-glass thermometers may be used to measure (1) environmental temperature and humidity, (2) internal temperature of animals and plants, (3) temperature of solutions involved in chemical and metabolic processes, (4) temperatures of water baths, autoclaves, incubators, and low temperature furnaces, (5) surface temperatures of interfaces, plants, animals, and man (measurement often being facilitated by rolling the thermometer over the surface), (6) maximum and minimum temperatures with thermometers of the self-registering type, and (7) differential temperature increments with the Beckmann thermometer. The liquid-in-glass thermometers also serve as laboratory standards for calibrating electrical thermometers.

4. Limitations of Liquid-in-Glass Thermometers

From the point of view of biophysical research methods, the most serious limitation of the liquid-in-glass thermometer is its lack of sensitivity. Temperature changes of 0.001°C. can be measured with the Beckmann type of thermometer, but it is often desired to measure temperatures one-hundredth or less of this value. This limitation is due to the fact that the ratio of bulb volume to bore cannot be increased indefinitely and also to the fact that a sensitive thermometer requires a considerable amount of heat to produce temperature changes. A second important limitation is that it is often desirable to measure temperature changes in tissues without adding appreciable

thermal capacity to the system.   A sensitive thermometer is usually slow, requiring several seconds or even minutes to reach equilibrium. This delay is known as thermometric lag and is due to the fact that heat is exchanged between the thermometer and the surrounding medium at a rate proportional to the ratio of the thermal capacity of the thermometer to the area of the bulb exposed.   Formulas derived from differential equations are available for determining thermometric lag time.   If the thermal capacity of the medium is large compared to that of a thermometer and the rate of stirring is adequate, errors due to thermometric lag can be ignored.   However, lag time is often important in measurements of air temperature, particularly if there are rapid changes.   A third important limitation is that liquid-in-glass thermometers often cannot be read *in situ* so that artifices must be resorted to, such as constricting the bore of a clinical thermometer. A fourth limitation is the difficulty in making precise measurements of temperatures between the fixed points at which the thermometer has been calibrated and under circumstances that differ significantly from those prevailing during the calibration.   According to Busse (3) this type of thermometer reached its highest development for precise work some fifty years ago, when electrical thermometers were not readily available.   In that period, individual portions of the scale were calibrated with mercury threads.   Corrections were made for internal and external pressure, effects of changes in bulb volume (*i.e.*, shift in zero reading), and other factors, so that a reproducible scale from 0 to 100°C. could be obtained.   Months of fine workmanship were spent in making and calibrating a single thermometer of uniform bore.

C. THERMOELECTRIC THERMOMETERS

Seebeck, in 1882, discovered that in a closed circuit, a small electromotive force is set up between a pair of bimetallic junctions at different temperatures.   Thermal electromotive force values are small, of the order of 40 microvolts per degree Centigrade, and are proportional to the temperature difference between hot and cold junctions.   Bimetallic junctions are called thermocouples and are extremely useful in determining temperatures in biological work. Frequently, several thermocouples are connected in series and are then called thermopiles.   A third dissimilar metal may be used in the circuit without affecting the thermoelectric force, provided both ends

of this metal are at the same temperature.   Measuring devices such as galvanometers or potentiometers may thus be inserted in the circuit without the creation of large extraneous thermoelectric forces.

For precise work, it is necessary to have as much of the external measuring circuit of the same metal as possible and, for this purpose, special galvanometers with all-copper circuits are available for use with copper–constantan thermocouples.   It is also advisable to investigate the parts of the circuit likely to give rise to troublesome thermoelectric voltages.   This can be done by closing the circuit and placing the hand near switch contacts, soldered connections, etc. When extraneous sources of e.m.f.'s are discovered in this manner, resoldering of contacts after careful cleaning and scraping will usually eliminate the difficulty.

For taking temperatures of small masses, thermocouples are very satisfactory.   They may be constructed of metal plates or wire of any gage.   It is possible to make junctions of extremely fine wire or metallic films, with the aid of a dissecting microscope.   These are useful in measuring the temperatures of bacterial colonies, insects, plants, or temperatures within the various organs and blood vessels of animals and man.

Thermocouple wires should be thoroughly cleaned, and the ends twisted together and soldered with a minimum of solder.   No soldering flux should be used that might cause the couples to deteriorate or vary in thermoelectric properties with time.   After soldering, the wires should again be cleaned with an organic solvent such as alcohol, ether, or acetone.   The "reference" junction may then be dipped into enamel, Ambroid, or other material to form a thin waterproof coating. Another method of protecting this junction is to fill a glass capillary with melted wax and insert the junction so that it rests within the thin tip of the capillary.   The "reference" junction is then inserted into a vacuum bottle or Dewar flask to the same depth and relative position as the bulb of a calibrated mercury thermometer.   The thermocouple and the thermometer are supported by a cork or rubber stopper fitted into the container.   In no case should either of these touch the bottom or sides of the container.   Three additional holes in the stopper will provide for insertion of a funnel, a siphon provided with a pinchcock or clamp, and a stirring device.   The latter may be rotated by manipulation of a string looped about the shaft or by an electric motor.   If the stirring is too rapid, mechanical heat may be generated and alter the temperature within the reference flask.

With the cold junction at 0°C. and the hot junction at t°C. the two
constants, B and C, may be calculated by determining the thermo-
electric force for two different temperatures of the warm junction.
The corresponding temperature may then be calculated for any other
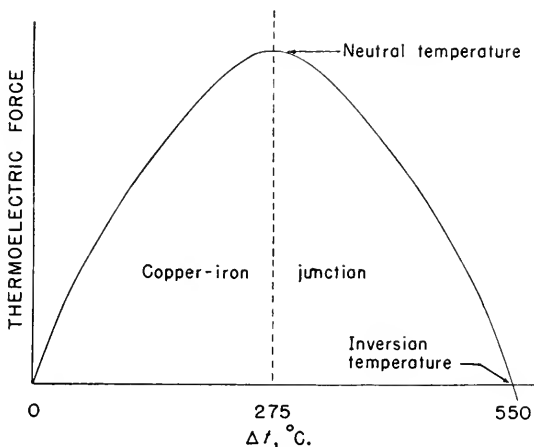observed thermoelectric force.

Fig. 2.   Relationship of thermoelectric force to tem-
perature difference between hot and cold junctions.

2. Measurement of Thermoelectric Force

Either of two fundamental methods may be employed in a variety
of ways to measure the thermoelectric force. These are: (a) the
galvanometric method, and (b) the potentiometric method.

Galvanometric Method.   This is the simpler and less accurate
of the two methods, the thermocouple leads being connected to
terminals of a galvanometer.   Sensitive pointer galvanometers are
now available, although for the most sensitive work lamp and scale
galvanometers are required.   In some instances, electronic or opti-
cal amplification can be employed and temperature changes as small
as 0.000001°C. can be detected.   The absolute limit of sensitivity
is determined by Brownian motion of the galvanometer coils or the
Johnson noise of the resistors in the electronic circuit.   For quantita-
tive measurements it is often convenient to arrange the amplification
so that a change of 1°C. corresponds to an even number of scale di-
visions.   For example, the amplifying system can be adjusted so

that 0.1°C. will be equal to 1 cm. of deflection. The accuracy of measurement made by these methods is limited due to the fact that the indicating instruments are not built to give linear displacement throughout their range. Furthermore, the temperature and humidity of the laboratory may be factors affecting the calibration of such devices.

The selection of the particular galvanometer system to be used will depend upon the particular experiments. Insensitive, rugged instruments of the pointer type can be used for large temperature differences between hot and cold junctions. In ordering a galvanometer for a specific purpose, it is important to specify the resistance of the thermopile in order that the galvanometer may be chosen to give approximately the correct damping. Most of the modern moving-coil instruments have magnetic shunts so that critical damping can be obtained for a large range of external resistances—i.e., 10 to 300 ohms. Careful thermal and electrical shielding are required for measurements of very small temperature changes. Introduction of the breaker type amplifier (the Perkin-Elmer Corporation) has made it possible to measure voltages conveniently as low as 5×10^{-9} v., corresponding to 0.0001°C.

Potentiometric Method. Whereas the galvanometric method gives a scale reading in direct proportion to the thermoelectric force flowing through the instrument, the potentiometric method uses a galvanometer or similar instrument as a null indicator for balancing the thermoelectric potential against a known potential in a second circuit, as shown in Figure 3.

The temperature range, sensitivity, and accuracy of the potentiometric circuit are determined by the relative value of the various resistances, the magnitude of the current supplied, and the sensitivity of the galvanometer. To obtain a standardized, reproducible current from the dry cell, the switch S_1 is thrown into position 1 and the current flowing through the galvanometer is adjusted until the voltage drop across a standard cell resistor, AB, equals the voltage supplied by the standard cell. Insertion and use of S_2 and a resistance, R, in the circuit is optional and serves to protect the standard cell. If switch S_2 is left open while coarse adjustment of current through the standard cell circuit is being made and then closed to short-circuit the resistance R during fine adjustment of the current, the standard cell circuit will not be in danger of passing large amounts of current. Switch S_1 is then thrown into position 2 and the voltage of the ther-

mocouple is determined by adjusting the slide wire contact until the
potential drop in resistor BC equals the potential of the thermocouple.
The slide wire is provided with a scale divided into suitable units.
The galvanometer serves only as a null indicator in this type of poten-
tiometer.

Potentiometers may be used with great accuracy. They may be
operated manually or they may be made self-recording. If the stand-
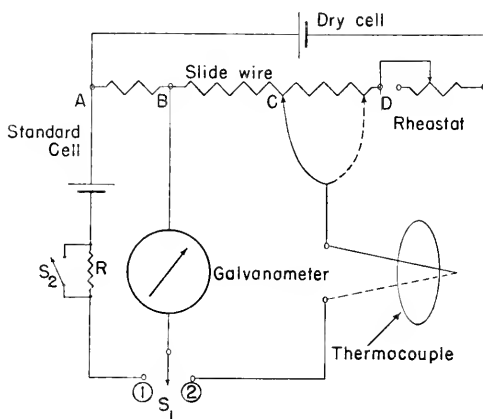ard junction temperature changes, the galvanometer must be ad-

Fig. 3.   Potentiometric circuit for thermocouples with standard cell.

justed. It is possible to compensate automatically for these tempera-
ture changes by installing the standard junction within the potenti-
ometer and attaching a bimetallic spring to the control spring of the
galvanometer coil.

A second type of potentiometric circuit for use with thermocouples
is shown in Figure 4 (4). Materials for the circuit are available in
most laboratories and it may be constructed readily at little expense.
It has the advantage of not requiring a standard cell. Basically, the
circuit employs a galvanometer as a null instrument to balance a
known current supplied by a 1.5 v. dry cell against the thermoelectric
current. The current is set at a constant value on the milliammeter
by adjustment of a 200 ohm (coarse) and a 20 ohm (fine) variable
resistor. The galvanometer is brought to the null point on each
temperature determination by means of a 7 ohm slidewire resistance
calibrated to read either in divisions or degrees of temperature. As

shown in Figure 4, leads to the rotary selector switch and the common lead to the potentiometric circuit are of copper (represented by solid lines). The constantan leads (dotted lines) are shared by the reference junction and all other junctions to be read against it. It is advisable to use a second reference junction inserted into the same vacuum bottle with the first and provided with a switch so that they
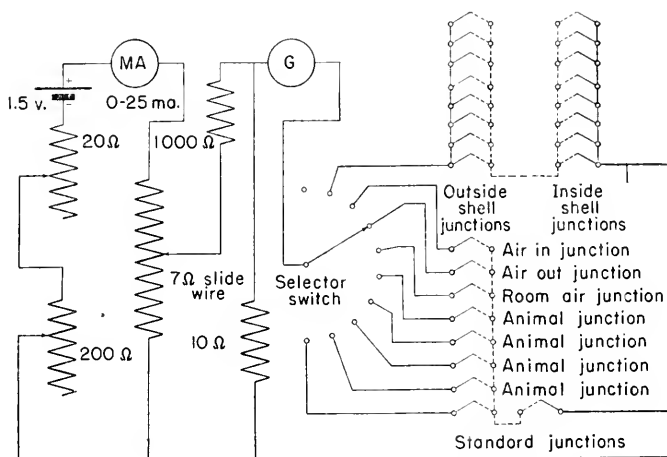
Fig. 4. Potentiometric circuit for thermocouples without standard cell.

may be read against each other to determine the zero point of the galvanometer. This will eliminate the effects of extraneous voltages from thermojunctions that inadvertently occur in the potentiometric circuit. One of the reference junctions is then used with the other thermocouples.

3. Calibration of Thermocouples

Tables published by the National Bureau of Standards are available relating the temperature difference between junctions of the most commonly used thermocouple metals and the voltage produced. If thermocouple wire has been purchased from a standard supply house (Driver-Harris or Leeds and Northrup), the tables published can be used as a satisfactory calibration. If wire from an unknown source is used, the thermocouples should be calibrated by comparison with a mercury thermometer, keeping one junction at constant temperature, e.g., in an ice bath.

Temperature of the cold junction as well as that of the warm junction affects the thermoelectric force developed. If the thermocouple is calibrated with the cold junction at $t_0°$C. and is used in an experiment at $t_1°$C., it may be necessary to add the value $(t_1 - t_0)k$ to the observed temperature. The constant, k, varies with the bimetallic junction used and the hot and cold junction temperatures. Fortunately, k (slope of the nearly straight portion of the curve in Fig. 2) is 1.0 for differences of 0 to 50°C. between copper–constantan junctions, 0 to 100°C. between iron–constantan and chromel–alumel junctions. Thus, a nearly linear relationship exists between temperature and electromotive force for these junctions in the ranges given and the observed electromotive force can be considered as proportional to the Δt of hot and cold junctions without use of correction factors. For precise measurements, even over a narrow range, the parabolic nature of the curve must be taken into consideration.

Thermocouples may be connected in either series or parallel as shown in Figure 5. Greater sensitivity can be obtained from thermocouples connected in series (thermopile) as the total voltage increases in proportion to the number of thermocouples. For certain purposes, parallel connection is satisfactory since the voltage produced by the thermopile is then approximately the average of that of the individual components.
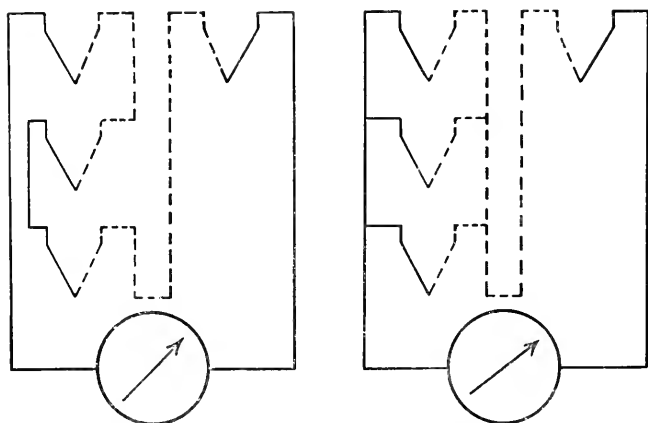
Fig. 5. Two thermocouple circuits. *Left*, series circuit; *right*, parallel circuit. Cold junction on right of each circuit, warm junctions on left.

4. Applications of Thermocouples

Although thermocouples can be used wherever liquid-in-glass thermometers are applicable, they have the additional advantages of small heat capacity and they can be read *in situ* by remote connection. This is a particularly important consideration when it is desired to measure the temperature of a patient in a fever cabinet or during cryotherapy (5). Under these circumstances, it is important that the patient's temperature be recorded from minute to minute. Furthermore, thermocouples have more than 100 times the sensitivity of a liquid-in-glass thermometer and are ten times as precise in the hands of the ordinary observer.

Particular applications of thermocouples to the measurement of internal temperature may be described to demonstrate the great flexibility of this method. Some of these are as follows:

Rectal Thermometers. A simple type of rectal thermometer was described by Hardy and Soderstrom (6) and consists of a single thermocouple and a galvanometer. The cold junction is immersed in a Thermos flask in warm water within 3°C. of body temperature.
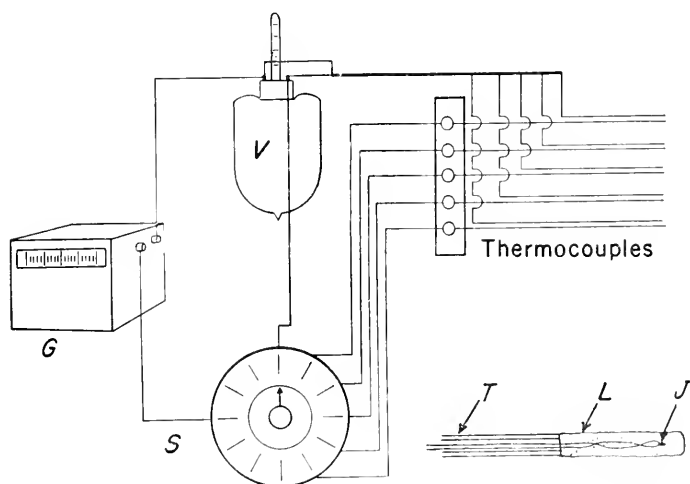
Fig. 6. Circuit diagram of rectal thermometers: (*G*) galvanometer; (*V*) vacuum bottle; (*S*) rotary switch. *Lower right*, construction of rectal thermometer: (*T*) rubber tubing; (*L*) latex coating; (*J*) thermal junction. After Hardy, Duerschner, and Muschenheim (7).

The rectal thermometer consists of a silver tube 25 mm. long, 3 mm. in diameter and 1 mm. in wall thickness. A small bead of Wood's metal is dropped into the bottom of the silver tube and the thermojunction embedded within it. The leads from the thermocouple are carried through small rubber tubing. The instrument is particularly adapted for use in children and laboratory animals. Readings are made directly from the galvanometer scale which is calibrated in degrees Centigrade. To make a reading, the thermometer is inserted into the Thermos flask with the cold junction and the galvanometer set to read the flask temperature. The thermometer is then removed and inserted approximately two inches into the rectum. Reliable readings can be obtained to $\pm$ 0.02°C.

For smaller laboratory animals such as mice or guinea pigs, the junction measuring rectal temperature is embedded in latex rather than contained in a silver tube. This prevents rupturing the colon and permits retention in the animal for periods as long as 48 hours without damage to the rectal mucosa. Figure 6 shows the arrangement for quickly measuring the temperature of as many as ten animals (7). More elaborate rectal thermometers are available commercially. These instruments are usually of the recording potentiometer variety and have a precision of ± 0.01°C.

Blood and Tissue Temperatures. Recently, Bazett (8) and co-workers have described a method for obtaining measurements of arterial and venous blood temperature by inserting a catheter containing a fine thermocouple into arteries and veins for considerable distances. This technique opens a new field for the measurement of blood flow in various parts of the body and for investigating more thoroughly the temperature-regulating functions of the hypothalamus. Since these catheters can be inserted without undue hazard to the subject, studies of the thermal exchanges in the larger organs of the body can now be made under a wide variety of circumstances.

Thermocouples inserted in long hypodermic needles are available on the market for the measurement of intramuscular temperature. More accurate measurements of intramuscular temperature can be made by sewing into the muscle fine wire thermocouples suitably protected with insulating varnish. *In vivo* studies of the heat of muscular contraction in man and animals have been possible with these techniques. Surprisingly wide variations in temperature have been dis-

covered in the extremities of resting man—particularly in the regions of large blood vessels (9). Hill and co-workers (10-12) have used thermopiles for measuring the heat of nerve conduction and of muscular contraction and have studied the effects of many variables upon this heat. Extreme sensitivity is required for these measurements since the temperature changes are of the order of 0.0001°C.
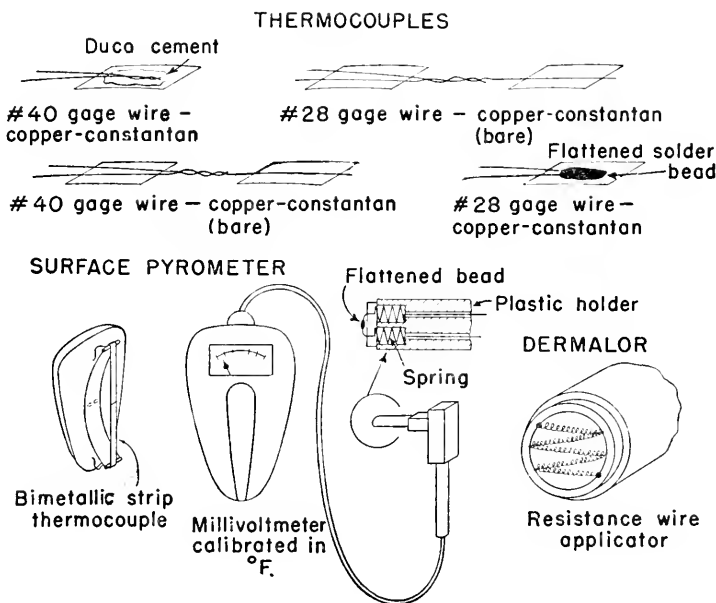

Fig. 7. Surface temperature measuring devices. Methods of mounting thermocouples shown in upper portion of figure, surface pyrometers in lower portion. After Stoll and Hardy (18). See text page 152.

The small dimensions of the thermocouple permit very rapid recording. Response times as short as one millisecond can be obtained with fine wire thermocouples. Using thermocouples of this type Adrian (13) and Bronk (14) have been able to record the rise in temperature due to the passage of an impulse down a nerve fiber.

For studies of the type just mentioned, no other thermometer has the flexibility, sensitivity, and accuracy that is necessary. Thus wherever such temperature measurements are anticipated, the ability to make thermocouples in the laboratory is desirable. For precision

thermoelectric thermometry, the reader is referred to the classical papers of W. P. White (*15–17*).

Surface Temperatures. One of the most difficult problems from a technical standpoint is that of measuring surface temperatures. The thermocouple is the most widely used thermometer for this purpose. There has been no standard procedure for mounting the thermocouple on the test surface and it has been tacitly assumed that simple contact of the thermocouple wire with the surface to be measured is sufficient. Figure 7 shows several of the methods that have been used for mounting surface thermocouples (*18*). Other

TABLE I

Performance of Various Surface Measuring Devices under Experimental
Conditions (after Stoll and Hardy, *18*)

Instrument	Performance under experimental conditions				
	Room (normal)	Wind velocity 4 ft./sec.	Infrared radiation	1500 W. lamp radiation	1500 W. lamp wind velocity 2 ft./sec.
Dermal radiometer.........	Excellent ±0.05 °C.	Excellent	Excellent	Requires correction	Requires correction
Thermocouple #40 gage wire (bare)........	Excellent	Fair	Good	Good	Excellent
Thermocouple #28 gage wire (bare)........	Good ±0.15 °C.	Fair	Poor	Poor	Poor
Thermocouple solder bead (adhesive tape)..	Good	Excellent	Fair	Poor	Fair
Thermocouple #40 gage wire (glued).......	Fair ±0.30 °C.	Fair	Excellent	Fair	Fair
Dermalor resistance thermometer..........	Fair	Good	Poor ±0.90 °C.	Poor	Good
Pyrometer strip thermocouple	Very poor ±3.0 °C.	Very poor	Very poor	Very poor	Very poor
Pyrometer solder bead.........	Very poor	Very poor	Very poor	Very poor	Very poor

methods should be mentioned such as sewing thermocouples into a cotton material stretched over the surface, and attaching the thermocouples to a copper screen wire brought into contact with the surface (*19*). Since the thermocouple will only measure its own temperature, it is apparent that it may not be measuring the true surface temperature when mounted in any of the ways mentioned above. Stoll and Hardy (*18*) have made a partial analysis of the thermocouple method of measuring surface temperature and, as is seen in Table 1, thermocouple thermometers vary considerably depending upon the

Fig. 8. Application of thermocouples to small animal calorimetry. Double-walled gradient calorimeter with thermocouples soldered directly to copper walls. (A) Outer shell, (B) inner shell, (C) wire animal cage, (D) thermocouple lead for ingoing air thermometer, (E) outgoing air thermocouple, (F) pan filled with oil for animal excreta, (G) shell thermocouple leads, (H) inner top showing window, thermocouple leads, and rim for airtight seal to inner shell, (I) outer top with thermocouple leads.

Animal Calorimeter
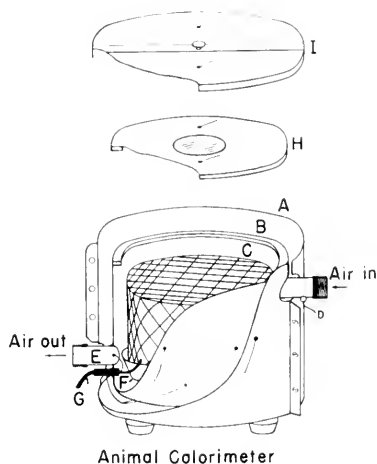
circumstances under which they are used. It is safe to say that dependable measurements of surface temperature cannot be made with thermocouples under all the conditions of the biophysical laboratory and the hospital. This inherent difficulty applies to the measurement of leaf temperatures, wall temperatures, and interface temperatures generally. However, circumstances may be such that the approximation obtained by the thermocouple may be the only practical one, as, for example, the measurement of the skin temperature of man under the clothing.

Calorimetry. Thermocouples are indispensable for calorimetry studies because of their high sensitivity and low thermal capacity. Recently described gradient calorimeters (*4,20*) depend entirely on the sensitivity and accuracy of thermocouples. In the second calorimeter, the measurement of heat loss depends upon the determination of the temperature difference between two copper cylinders, the inner

of which contains the animal under study (see Fig. 8). The thermo-couple arrangement for measuring the temperature difference be-tween the cylinders, temperature of the room, the skin and rectal temperature of the experimental animal, and air temperature within the chamber is shown in Figure 4. This type of calorimeter, because of the small heat capacity of thermocouple arrangements, permits measurements of heat loss within periods of ten minutes. A very recent modification of this technique described by Benzinger (21) attempts to achieve a total response time of less than one minute.

5. Limitations of Thermocouples

Thermocouples yield less precise temperature measurements than resistance thermometers. This is not serious for most biophysical work because the lack of precision is insignificant within the range of most biological temperatures.

Relative to the liquid-in-glass thermometers, the thermocouple thermometers are more expensive and more difficult to manipulate. For the extremes in sensitivity, a delicate galvanometer system is re-quired. The apparatus is not portable and, for great precision, the best potentiometers and experienced observers are required (3). Base metal thermopiles, which are commonly used for biophysical work, have a tendency to change slightly with age. This requires troublesome recalibration.

D. RESISTANCE THERMOMETERS

Resistance varies with temperature according to the approximate experimental law:

$$R_t = R_o(1 + \alpha_0 t) \tag{4}$$

where R_t and R_o are resistance at temperatures t and $0°C.$, respec-tively, and α_0 is the temperature coefficient of resistance referred to $0°C.$ The resistance of insulators, electrolytes, and carbon decreases as temperature rises and is therefore negative. Metals have a posi-tive coefficient, the value of which is approximately 0.0038 per degree Centigrade for a large number of pure metals. The relationship be-tween temperature and resistance can be applied effectively for three types of resistance thermometers useful in biophysics. These are: (1) noble metal resistance thermometers, (2) base metal resistance

thermometers, and (3) thermistor or semiconductor resistance ther-
mometers.

All three types consist of a resistance element and an electrical
circuit for measuring resistance changes with temperature. For
temperatures above 0°C., Callendar (22) formulated the relationship
between temperature and resistance as:

$$t = \left(\frac{R_t - R_o}{R_{100} - R_o}\right) 100 + \delta \left(\frac{t}{100} - 1\right) \frac{t}{100} \tag{5}$$

where t is the temperature in degrees Centigrade, R_t is the resistance
of the resistor at temperature t, R_o and R_{100} are its resistance at 0° and
100° C., respectively, and δ is a constant characteristic of the indi-
vidual thermometer and must be determined by calibration at a fixed
point. The value of δ usually lies between 1.49 and 1.50°C.

1. Types of Resistance Thermometers

Noble Metal Resistance Thermometers. Platinum wire is
usually employed in noble metal thermometers. This metal is
readily obtainable in a high degree of purity. Its purity is assured
if the mean temperature coefficient, α, (equation 4) between 0 and
100°C. is not less than 0.00388 and the constant δ of the Callendar
equation (5) is not greater than 1.52. The resistor of the thermometer
is usually supported in a framework so that it is protected from strain
and injury. It is enclosed in a protecting tubing or sheath and con-
nected by insulated copper wires to resistance measuring apparatus.
A resistor of this character requires about 2 m. of wire. The wire is
wound into a coil around a frame so that it can expand or contract on
heating, with a minimum of mechanical stress and time lag. The
heating effect of the measuring current should be minimal. Mica
can be used for the mounting frame, gold for the connecting leads in-
side the protecting tube, which may be of metal, glass, porcelain or
fused silica, and copper wire for the leads to the measuring instru-
ment.

Base Metal Resistance Thermometers. Nickel, Nichrome, or
similar alloys are most commonly used in base metal resistance
thermometers. A large number of different alloys with a wide
range of temperature coefficients are available. Selection of a par-
ticular alloy will depend upon cost and the particular resistance–

temperature characteristic that fits the application for which the thermometer is to be used.

Semiconductor Resistance Thermometers (Thermistors). Oxides of manganese, nickel, cobalt, copper, and uranium belong to the class of chemical compounds known as semiconductors which are extremely sensitive to relatively minute temperature changes. A new series of resistance units known as "thermistors" has been developed by the Bell Telephone Laboratories and may further ex-
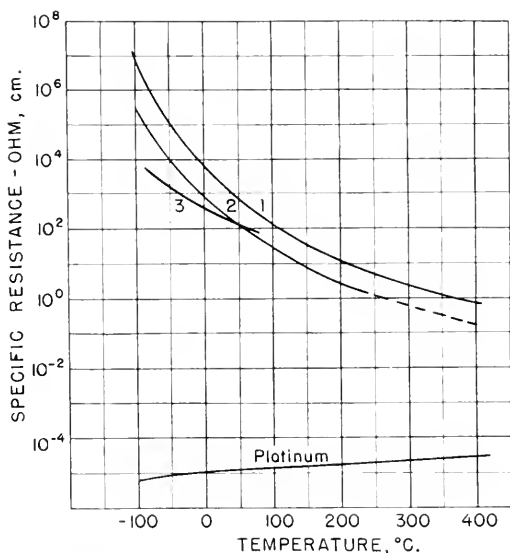

Fig. 9. Resistance–temperature relationships of three thermistor materials compared to platinum. (Courtesy *Bell Telephone Laboratories.*)

tend the use of resistance thermometry for biophysical temperature determination (23). Essentially, a thermistor consists of an oxide or mixture of oxides with a high temperature coefficient of resistivity (Fig. 9), a method of mounting and protecting the oxides, and electrical leads. Over thirty variations in the prototype are now available. All are small, inexpensive, flexible in application, and undergo practically no change in resistance properties with continual use. They are available in three basic shapes (Fig. 10)—discs, rods, and beads. Various mixtures of the oxides can be used to provide the resistance-temperature relationships desired for the particular application.

The oxides are usually enclosed within ceramic or glass fired under accurately controlled temperature and atmospheric conditions. Lead wires are attached and the resistance unit enclosed or mounted in a manner suitable to the thermal measuring requirements. Stability
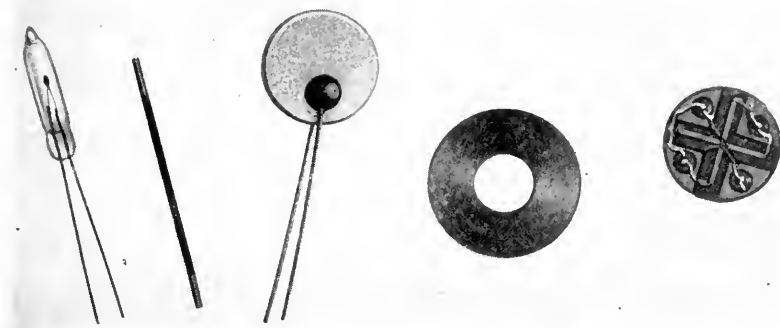
Fig. 10.   Thermistors made in the form of a bead, rod, disc, washer, and flakes.
(Courtesy *Bell Telephone Laboratories*.)

tests on thermistor units have disclosed no appreciable change in resistance characteristics through 500,000 heating and cooling cycles.

2. Electrical Circuits for Resistance Thermometers

Figure 11 shows a diagrammatic presentation of the most commonly used circuit for resistance thermometers of the three lead type. The galvanometer is used in this circuit as a null indicator to show when the bridge is in balance. Current flows through the galvanometer only when electrical balance is not present. This method is independent of voltage surges from the current source and a dry cell or low voltage alternating current with a copper oxide or selenium rectifier can be satisfactorily used as this source.

A double slide wire Wheatstone bridge is represented in the diagram. Accuracy of measurement is assured by this circuit design, which places all moving contacts in the galvanometer or battery circuits. Both slide wires are mounted on a single drum and have resistance values such that, at all points on the slide wire, the resistances of arms 1 and 2 are equal so that the ratio arms of the bridge will be equal in resistance. Resistance AB will vary as the dial is rotated and since arm 1 equals arm 2, the bridge will be balanced when arm 3

equals arm 4. This condition is obtained when the variable resist-
ance of AB equals the varying resistance of the detector coil, BT.
Direct temperature readings at T (rather than ohms) are possible
with proper calibration of the slide wire, which constitutes the vari-
able section of the resistance of arm 3.

Although the above circuit is the most generally satisfactory for
ordinary biophysical work, precision resistance thermometry em-
ploying a platinum wire thermometer requires somewhat different
circuits. The basic operation of precision resistance thermometry
consists in the measurement of the potential drop across the resistance
thermometer as compared to that across a standard resistor. Either
a potentiometer or a bridge circuit can be used. The potentiometer
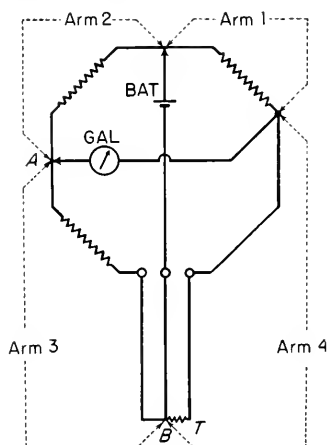
Fig. 11.   Commonly used Wheat-
stone bridge circuit for three lead re-
sistance thermometer (modeled after
a Leeds and Northrup circuit).

is a more versatile instrument capable of measuring temperatures
over a greater range than a bridge circuit with equal sensitivity, but
is somewhat more difficult to use. In the potentiometric method
the voltage drop across two resistors carrying the same current is
measured. One of these resistors is a standard of the four terminal
type and the other is the resistance thermometer. Both are con-
nected in series with a battery. Resistance thermometers for use in
the biophysical temperature range usually have a potential drop of
0.1 v. or less and are best used with low voltage potentiometers.
Mueller (24) states that for accuracy of 0.001°C. in temperature de-
termination, resistance measurements must be made with an accu-
racy of two to four parts in a million.

The Wheatstone bridge circuit (25)—as used with the four lead
platinum resistance thermometer—is shown in Figure 12. In this
circuit:

$$R_1/R_2 = R_3/R_x \qquad (6)$$

Two observations are required with this circuit to measure the resist-
ance of the terminal element. R_1 and R_2 are the resistances of the
equal ratio arms. R_3 is the resistance required to balance the bridge.
R_x is the resistance between the branch points of the thermometer,
the four leads of which are designated as A, B, C, and D. By means
of the slide wire the ratio of the arms containing R_3 and R_x may be
adjusted to equality within one or two parts in ten million. Once the
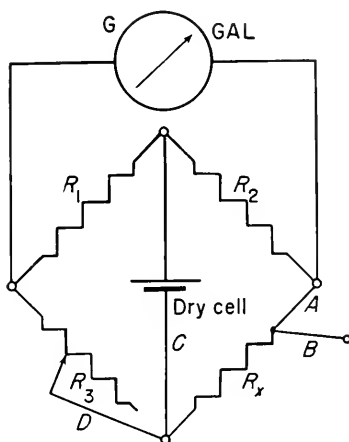
Fig. 12. Bridge circuit for four lead,
precision resistance thermometer.

bridge has been balanced, with the battery connected to thermometer
lead C, a commutator with amalgamated contacts is used to switch
the battery to lead B. The positions of thermometer leads A and D
are interchanged, and the bridge is again balanced. Since it can be
shown that:

$$R_x = (D + A)/2 \qquad (7)$$

the reversal of the leads will cancel out any effects of resistance in the
leads and the true resistance of the thermometer will be obtained.
A small variable resistor is sometimes placed in one of the thermom-
eter leads to equalize the lead resistance in the two positions of the
commutator. Unless it is desirable to determine temperature dif-

ferences smaller than 0.001°C., it is unnecessary to make tempera-
ture corrections for the bridge coils, the resistance of which varies
slightly with room temperature.

3. Calibration of Resistance Thermometers

Calibration of the thermometer at fixed points is laborious and for
most purposes it is sufficient to compare the resistance thermometer
to a calibrated mercury thermometer in a stirred bath, the tempera-
ture of which is varied slowly in steps over the desired range.   If
greater accuracy is desired, a precise platinum wire resistance ther-
mometer is used.   It is usually desirable to keep one of the latter for
use as a laboratory standard and to use base metal resistance ther-
mometers or thermistors—both of which are comparatively inexpen-
sive—for routine temperature measurements.

A graph may be made (as in Fig. 9) of ohms of thermometer re-
sistance against degrees Centigrade for the working temperature
range.   It is often feasible to calibrate the slide wire of the bridge or
potentiometer to give readings directly in degrees of temperature.

4. Applications of Resistance Thermometers

Platinum resistance thermometers in biophysics are used mostly
as laboratory standards.   Although they may be used over a wider
temperature range and are less susceptible to deterioration than base
metal thermometers, they require an expensive bridge or potentio-
metric circuit and more precise techniques if advantage is to be taken
of their greater precision.

At the present time, base metal resistance thermometers—usually
Nichrome—are used for almost all routine biophysical temperature
determinations in which resistance thermometers are employed.
Beside being inexpensive and easily obtainable, they give satisfactory
readings in the moderate temperature range and will hold a calibra-
tion within 0.02°C. for several years if suitably mounted and pro-
tected.

Some general applications of resistance thermometers are as fol-
lows: (a) Rectal thermometers that are rugged, yet very sensitive to
any significant changes in temperature, may be constructed (26)
(see Fig. 13).   (b) Temperature inside the clothing may be obtained
by threading resistance wire into a sewing machine and stitching the
resistance wire directly into the clothing (27).   (c) Resistance units

have been used successfully for many years in the walls of calorimeters and in other apparatus for the measurement of air and wall temperatures. (d) Resistance thermometers have often been used to measure surface temperatures. One type is the Dermalor of Figure 7 and a second type has been described by Soderstrom (26). Bohenkamp (28) measured human surface temperatures by winding resistance wire around various parts of the subject's body. (e) Resistance thermometers can be used to measure the temperature of plants, bacterial colonies, solutions, interfaces, and practically every place where liquid-in-glass or large thermocouples are used. Like thermocouples, they have the advantage over liquid thermometers of being read *in situ* either near to or remote from the observer. (f)
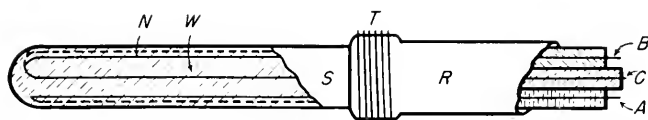
Fig. 13. Resistance type of rectal thermometer: (A) lead; (B) and (C) leads soldered together at tip of thermometer; (N) insulated nickel resistance wire wound around three leads; (S) silver shell; (R) rubber tubing covering leads; (T) thread binding rubber to silver shell. After Soderstrom (26).

Resistance thermometers are particularly adaptable for use with electrical recording apparatus and many direct reading and automatic recording temperature devices are available on the market.

Thermistors, in addition to many of the above applications, possibly can be used in rod form for intravascular temperature measurement (Fig. 10). The probe type is useful in measuring ambient air or liquid temperatures while the disc or pellet type can be easily inserted into a metal thermometer bulb. The disc form has also been sweated onto metal plates to give a low thermal impedance connection to the object whose temperature is being studied. The large disc form is enveloped in a paint finish for use in humid surroundings. Minute beads with response times of less than a second in air and flake thermistors with time constants from one millisecond to one second have also been manufactured. These forms are useful for temperature determinations of objects that are either inaccessible, in motion, or too hot for contact thermometry. Radiation from the object can be focused on the thermistor by means of a concave mirror.

5. Limitations of Resistance Thermometers

Resistance thermometers have the disadvantage of being much more difficult to construct than thermocouples for ordinary biophysical measurement. Also, it is not generally feasible to construct resistance thermometers of the small size attainable with thermocouples. This fact limits their use for the type of measurement requiring very small, sensitive elements as, for example, measurements of intravascular temperature. The introduction of the thermistor may permit such measurements although as yet they have not been so used. Resistance thermometers require more elaborate physical apparatus than any of the other temperature measuring methods described. For effective use, they also require a higher degree of manipulative skill. Resistance thermometers of the base metal variety may change in calibration with use. Soderstrom has reported that readings with his nickel resistance thermometers change as much as 0.2°C. during the first year of use and within 0.02°C. during the second.

E. TEMPERATURE MEASUREMENT BY RADIATION

1. General Principles

For high temperature measurements, optical pyrometers can be used. These instruments are essentially photometers employing monochromatic light—usually red—in which the intensity of radiation from a standard source is compared to that from the object being studied. The intensity is controlled by means of absorption filters or iris diaphragms interposed into the pathway of the radiation from either the standard source or the radiating object. These instruments are of interest in biophysics principally in the measurement of the heat energy output from furnaces, arc lamps, and other high temperature sources to which biological material may be exposed.

Of special interest to the biophysicist concerned with determining temperature of surfaces is the radiometer. The radiometer is an instrument that converts incident radiant energy into thermal energy, which can be measured quantitatively. The detecting element of a radiometer is usually either a thermopile or a resistance element (bolometer) connected to a suitable potentiometric or bridge circuit for measurement.

A crude radiometer can be devised by blackening the end of a

thermometer, enclosing a thermometer in a blackened sphere, or blackening a bulb filled with water, alcohol, or ether and noting in each case the heights to which the thermometric substance rises when radiation is allowed to impinge on the blackened surface. This type of radiometer is insensitive and slow, but because of ruggedness and simplicity has received considerable attention, as, for example, the Vernon globe thermometer.

The amount of energy received by a radiometer is often so small that electronic or optical amplification is necessary. This energy depends upon the difference in temperature between the radiometer and the surface being measured and upon the emissive power of the surface. The net transfer of heat by radiation can be written:

$$H_R = S_0 \epsilon_1 \epsilon_2 (T^4 - T_0^4) A \tag{8}$$

where H_R = heat transfer in g. cal./sec., S_0 = Stefan-Boltzmann constant = 1.37×10^{-12} g. cal./sec./cm.², T and T_0 = absolute temperature of hot object and its environment, ϵ_1 and ϵ_2 = emissivity of surfaces of radiator and environment with maximum values of unity, and A = radiating area of object.

A perfect black body has the highest emissive power or emissivity and is assigned the value of 1. All other objects have emissivities less than 1. Since the emissive power of an object is equal to the absorbing power of the object, emissivity can be related to the reflecting power (R) of an object with ϵ being equal to $1 - R$. This relationship provides a convenient means for measuring the emissivity of biologically important surfaces as, for example, human skin, leaves, animal fur, clothing, etc. Many objects may act as good radiators for one wavelength and as poor radiators for others. The human skin is a good example. It is important that the emissivity used in the radiation formula be that for the wavelength range in which the object is radiating. The emissivity for white human skin has a value of 0.99 in the infrared portion of the spectrum in which the skin radiates—although its emissivity in the visible spectrum would be as low as 0.3.

2. Calibration of Radiometers

The energy calibration for a radiometer can be most easily made with a radiation standard of the National Bureau of Standards. For low temperature work in the biological field, the calibration can be ac-

complished by means of two black bodies (Leslie cubes) at different temperatures. Two Leslie cubes are employed in order to rule out the necessity of measuring the temperature of the thermosensitive elements of the radiometer (*29,30*).

3. Applications of Radiometers

Radiometers have been extensively used in the measurement of skin and surface temperature. An instrument of this type is shown in Figure 14. The latter measurement is of particular importance
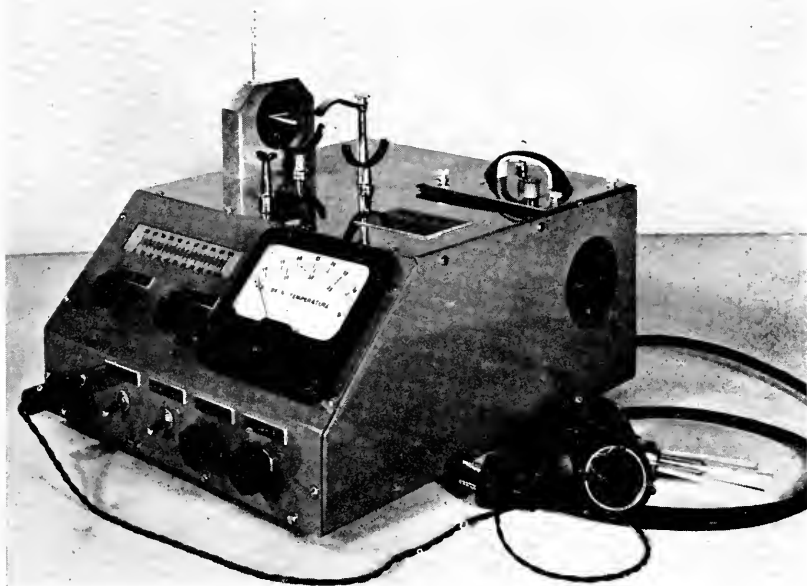

Fig. 14.   Hardy dermal radiometer.   (Courtesy *Baird Associates.*)

in bioclimatology. The radiant temperature of the nighttime sky, the terrestrial surroundings, and the sun are all of importance in evaluating thermal stress. Experimental investigations of the sensitivity of the human skin to temperature and to pain have involved the use of radiometric techniques as skin thermometers. As can be seen from Table I, there is no other instrument as dependable as the radiometer for temperature measurement of exposed surfaces.

A new type of radiometer for measuring the radiation temperature of the outdoor environment has recently been described by Richards and Hardy (*31*). This instrument consists of three spheres, black, white, and polished, into which electrical energy is fed to bring them all to the same temperature. This gives a measurement of radiation from which the "wall temperature" of the environment may be computed.

4. Limitations of Radiometers

The application of a radiometer is limited almost entirely to the measurement of surface temperatures in biophysics, and for this single purpose it is beyond question the best instrument. Radiometers are fragile instruments and require relatively expensive electrical equipment for measurement of radiation. Emissivity values must be worked out for each type of surface to be investigated and corrections often must be applied to the radiometric readings due to reflected and scattered light.

F. MISCELLANEOUS TEMPERATURE MEASURING DEVICES

All the following temperature measuring devices have some biophysical application although limited to some particular problem. The discussion is therefore brief.

1. Thermoluminescence

Many chemical compounds emit fluorescent radiation within a narrow temperature range only. Zinc sulfide produces yellow fluorescence between 0 and 123°C. The fluorescence of zinc oxide changes from red to green at 704°C. The quality and quantity of fluorescence of the microorganisms found in the slime forming on certain decaying fish changes to a marked degree as the environmental temperature is altered. Many other substances have other characteristic fluorescent temperatures and, if these critical points happen to fall near those of some biophysical reaction, they may be useful indicators.

2. Dielectric Constant Changes

The dielectric constant of most compounds is a function of temperature. A ceramic capacitor such as titanium dioxide, TiO_2, for

example, undergoes marked changes. A common laboratory substance, Rochelle salt crystals, exhibits a large shift in dielectric constant between 20 and 30°C. but is destroyed by temperatures over 105°. It is somewhat difficult to use these substances for temperature determination because resistance losses are large enough to require separate balancing for a null reading if a bridge circuit is employed to measure the change in capacitance. The dielectric constant shift can be used to better advantage to measure temperature if these substances are used as the capacitance part of an inductance–capacitance circuit or phase shift oscillator.

3. Thermosensitive Magnetic Alloys

Thermosensitive alloys of silicon steel, chromium, and nickel can be made with Curie points (T_c, point at which magnetization drops to zero) below 100°C. with rapid decrease in magnetization starting at a temperature value 60% of the Curie point. A coil using a core of one of these alloys can be employed with an audio-frequency circuit to measure temperature within the range of 60% (T/T_c) of the Curie point of that alloy.

4. Thermal Conductivity Cells

Temperature–resistance relationships in wires heated by a constant current or a constant terminal voltage and placed in glass cells are used to measure pressure, humidity (electric hygrometers), radiation, and temperature and to analyze gases. By determination of the rate of heat dissipation from the wire to the surrounding medium, a coil of fine platinum or tungsten wire may be used as a flow meter for either liquids or gases.

The thermoconductivity principle is also frequently applied in studies of human and animal physiology. The rate of dissipation of the heat in the intestine, for instance, may be determined by resistance changes in a heated wire enclosed within the walls of a balloon placed in the intestine (32). Similarly, a heated wire lead into a blood vessel through a hypodermic needle may be used to yield accurate determinations of the rate of flow of blood.

5. Bimetallic Strip Thermometers

Laminated strips of metals that have different temperature coefficients of linear expansion are widely used to determine tem-

perature.   Unless specially designed, however, their accuracy usually is less than 0.1°C.   Their use is limited to panel, wall, remote position, or other indicating thermometers of the dial type.   A combination of a metal with a high temperature coefficient, such as brass, and one with a low coefficient, such as Invar, gives the best result.

Bimetallic strips find their widest application as thermostats for control of constant temperature cells, baths, or rooms.   The metals selected should have low mechanical hysteresis and high tensile strength so that, when cooled or heated repeatedly, they will warp to the same degree consistently.   Such strips may be used to open or close electrical contacts or a mechanical valve, or to move indicators. If more exact temperature changes are required, the strip is used unloaded and its movements are followed by means of a wire strain gage or a change in capacitance between the bending strip and a fixed plate, or the movements can be followed by means of a phototube optical system.

6. Thermoscopes

The thermoscopic method is the simplest for measuring temperatures and is employed in the laboratory to designate a certain arbitrary temperature, usually for calibration of a calorimeter or electric furnace.   One form of thermoscope employs a small piece of lead, paraffin, or sulfur placed so that the melting point of any one of these or similar substances indicates the temperature of another mass under similar thermal conditions.   A hollow metal sphere that would pass through a given aperture at a certain temperature but not at any other is another form of thermoscope.   If the substance used as a thermoscope undergoes a change of state at some known "fixed point" (such as the melting or boiling point of sulfur) this fixed point can be related to a known temperature scale.   Although thermoscopes have the advantage of simplicity, they have the disadvantage of indicating only one temperature point and this only momentarily.

G. SPECIAL TEMPERATURE PROBLEMS IN BIOPHYSICS

Temperature is probably the most important universal factor affecting all life processes.   Biophysical research is concerned not only with temperature determinations on man but on the organisms sharing his environment and temperature factors in the environment itself.   The general methods available have been described in the pre-

ceding section. They can be modified to fit almost any biophysical requirement ranging from quantitative studies of photosynthesis in plants by highly sensitive, quick-responding thermocouples to determination of the rectal temperatures of an elephant. Chemical reaction rates are affected to an even greater degree by temperature than purely physical reactions such as conduction, convection, and diffusion. Photochemical reactions are affected very little. Reactions involving ionic exchange and neutralization of acid and alkali are so extremely rapid that temperature changes are immeasurable. In those reactions that proceed with measurable speed, the importance of temperature received added emphasis by the publication of the van't Hoff relationship (1884). It was shown that the velocity of a chemical reaction is at least doubled by a $10°C$. rise in temperature. This discovery led to the use of the "temperature coefficient" (Q_{10}) to express the relation of the reaction velocity at a given temperature to that at $10°C$. lower. It became apparent that the van't Hoff relationship might be used to decide whether certain physiological processes such as the conduction of a nerve impulse had a physical or chemical basis. An attempt by Snyder (1908) to decide the nerve conduction problem was inconclusive because the Q_{10} value of approximately 2 he obtained is borderline between physical and chemical reactions.

In determining temperature effects on the hydrolysis rate of sucrose, Arrhenius (1889) (33) broadened the van't Hoff relationship by expressing it as the differential equation:

$$d(\log k)/dt = \mu/RT^2 \qquad (9)$$

where k is the reaction velocity constant, R the gas constant $(= 1.98$ or roughly 2 cal.), and μ is a term that Arrhenius at first thought had no physical meaning but that was later shown experimentally to be a constant of thermodynamic significance. The term μ became known as the "temperature characteristic" of a reaction. According to the kinetic theory, the velocity of any chemical reaction is governed by the number of effective molecular collisions in a given time. In order for molecules to react, they must collide with relatively great energy (energy of activation). The temperature characteristic, μ, of any chemical reaction is defined as the energy required to "activate" the molecules entering into the reaction and is expressed in calories per gram molecular weight of reactant.

According to Crozier (34), the temperature characteristic may represent the critical increment of energy for the formation of active

molecules of a catalyst. The catalysts repeatedly involved in bio-
chemical reactions might thus be identified by determination of tem-
perature characteristics. Although many complex systems, involv-
ing serial reactions, have definite temperature characteristics, it is
seldom possible to assign the μ value to a particular reaction in the
system.

By rearranging the above equation, the temperature characteristic
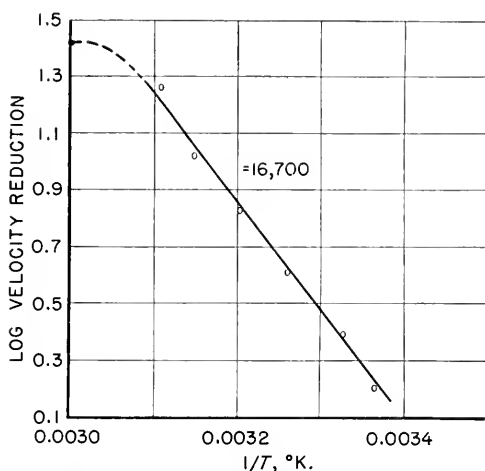of a reaction can be determined by measurement of reaction rates at

Fig. 15. Typical Arrhenius graph. Reciprocal of absolute tem-
perature plotted against logarithm of velocity of anaerobic reduction
of methylene blue by bacteria in presence of sodium succinate as
source of hydrogen. Destructive effects evident above 50 °C. Tem-
perature characteristic, μ, is 16,700. After Crozier; curve based on
data of Quastel and Whetham.

two known temperatures. It is more practical, however, to deter-
mine μ by plotting log k against $1/T$ for a series of observations at
several temperatures. The temperature characteristic will then be
the slope of the line obtained as shown in Figure 15. The temperature
characteristic has been determined for many physiological processes,
including cell division rates, rate of pupal development, heart rate,
velocity of ciliary movement, velocity of the nerve impulse, growth
rate, carbon dioxide production and oxygen consumption rates in
many plant and animal forms *in vivo*. It has also been determined for
many purely chemical or enzymic processes *in vitro*, such as the inver-

sion of sucrose by acids or invertase, conversion of fat to soap by alkali, and many oxidative reactions.

Many physiological processes involve several complex chemical reactions and the determination of the dominant reaction is the chief application of the temperature characteristic. The usual procedure is to study the reaction in a bath, the temperature of which is either constant or variable at a controlled rate.

Theory of the Master Reaction

Crozier (*34,35*) proposed that, in a complicated physiological reaction such as cellular oxidation, there may be a slowest step which governs the over-all velocity of the process. He further suggested that in this type of reaction, the temperature coefficient of the reaction velocity (Q_{10}) yielded a quantity similar to energy of activation. This quantity (the temperature characteristic or μ) changed in value on passing from one temperature range to another and these changes were interpreted as shifts in "mastery" from a reaction with a velocity constant $k_1(\mu_1)$ to a reaction with a velocity constant $k_2(\mu_2)$. Crozier believed that k_1 and k_2 were constants for the rates of formation and destruction of an enzyme. By dominating the steady state concentration of an enzyme, these constants also regulated the over-all velocity of the reaction. Many enzymes are common to many different biological species and different life processes. Thus, it might be predicted that a random distribution of temperature characteristics for many diverse physiological reactions in different species would not exist. Rather, well defined modes corresponding to either k_1 or k_2 for various enzymes would be expected. Crozier's plot of μ values against frequency of their occurrence shows these modes.

The rate of the heart beat might be determined, for example, by a single type of activation because of the interdependence of the several reactions involved (*35*). In any such related chain of reactions, it would be the slowest reaction that would determine the rate for the entire process and act as the "master reaction." The chain of reactions associated with muscular contraction, flavoprotein systems, and cytochrome, are also examples. In these cases in which the Arrhenius plot does not give a straight line, Crozier calls the point of slope change the critical temperature at which dominancy changes from one master reaction to another.

Considerable experimental evidence has been collected by Crozier

and others in support of the "master reaction" theory.   Burton (*36*)
has been the chief critic of the theory, basing his objections on the
law of mass action.   He believes that if one reaction is to dominate
over another in the same reaction chain at one temperature but may
in turn be dominated by another reaction at another temperature, the
temperature characteristic of the two reactions would have to vary by
a larger interval than that actually observed.   By mathematical
analysis, Burton proposes that the slope of the line of an Arrhenius
plot may result from pace-setting shared by several reactions of dif-
ferent critical increments rather than control by a master reaction.
Both Burton's criticism and Crozier's original suggestions fail to
offer any explanation of an Arrhenius plot with regard to underlying
mechanisms.   Johnson, Eyring, and co-workers have extended the
work of other investigators to develop the theory of rate processes.
Additional information on this subject may be found in the book by
Glasstone, Laidler, and Eyring (*37*).

References

1. Waidner, C. W., E. F. Mueller, and P. D. Foote, *Symposium on Py-
      rometry*, American Institute of Mining and Metallurgical Engineers,
      1920, p. 46.
2. Busse, J., "Thermometry," in O. Glasser, *Medical Physics*.   Year Book
      Publishers, Chicago, 1944.
3. Busse, J., "Liquid-in-Glass Thermometers," in *Temperature, Its Meas-
      urement and Control in Science and Industry*.   Reinhold, New York, 1941.
4. Prouty, L. R., M. J. Barrett, and J. D. Hardy, *Federation Proc.*, **7**, 96
      (1948); and other material to be published (a simple calorimeter for
      simultaneous determination of heat production and heat loss in labora-
      tory animals).
5. Fay, T., W. Bierman, and M. Friedlander, *Arch. Phys. Therapy*, **21**, 585
      (1941) (the penetrative effect of cold).
6. Hardy, J. D., and G. F. Soderstrom, *Rev. Sci. Instruments*, **8**, 419 (1937)
      (apparatus for surface and body temperature).
7. Hardy, J. D., D. R. Duersehner, and C. Muschenheim, *J. Infectious Dis-
      eases*, **72**, 179 (1943).
8. Bazett, H. C., and B. McGlone, *Am. J. Physiol.*, **82**, 415 (1927) (temper-
      ature gradients in tissues).
9. Sheard, C., G. Roth, and B. Horton, "Normal Vasoconstriction, Vaso-
      spasm and Environmental Temperature," and C. Sheard, M. Williams,
      and B. T. Horton, "Skin Temperature of the Extremities Under Vari-
      ous Environmental and Physiological Conditions," in *Temperature*,

Its Measurement and Control in Science and Industry. Reinhold, New York, 1941.

10. Downing, A. C., R. W. Gerard, and A. V. Hill, *Proc. Roy. Soc. London,* **B100,** 223 (1926).
11. Hill, A. V., *Adventures in Biophysics.* Univ. Pennsylvania Press, Philadelphia, 1931.
12. Hill, A. V., *Muscular Activity.* Williams & Wilkins, Baltimore, 1926.
13. Adrian, E. D., *The Basis of Sensation.* Cristophers, London, 1928.
14. Bronk, D. W., *J. Physiol.,* **71,** 136 (1931).
15. White, W. P., *J. Am. Chem. Soc.,* **36,** 1856 (1914) (thermoelectric installations, especially for calorimetry).
16. White, W. P., *The Modern Calorimeter.* Chemical Catalog Co., New York, 1928.
17. White, W. P., *J. Am. Chem. Soc.,* **36,** 2292 (1914) (thermoelements of precision).
18. Stoll, A. M., and J. D. Hardy, *Federation Proc.,* **7,** 120 (1948) (direct experimental comparisons of several surface temperature measuring devices).
19. Palmes, E. D., and C. R. Park, "An Improved Mounting for Thermocouples for Measurement of the Surface Temperature of the Body," Medical Department Field Research Laboratory Publication, Fort Knox, Kentucky, March, 1947.
20. Day, R., and J. D. Hardy, *Am. J. Diseases Children,* **63,** 1086 (1942) (calorimeter for measuring heat loss of premature infants).
21. Benzinger, T., personal communication.
22. Callendar, H. L., *Phil. Trans. Roy. Soc. (London),* **178,** 160 (1887).
23. Becker, J. A., C. B. Green, and G. L. Pearson, *Bell System Tech. J.,* **26,** 170 (1947) (properties and uses of thermistors).
24. Mueller, E. F., "Precision Resistance Thermometry," in *Temperature, Its Measurement and Control in Science and Industry.* Reinhold, New York, 1941.
25. Smith, F. E., *Phil. Mag.,* **24,** 541 (1912) (bridge methods for resistance measurements of high precision in platinum thermometry).
26. Soderstrom, G. F., *Rev. Sci. Instruments,* **4,** 285 (1933) (resistance thermometers).
27. Burton, A. C., *J. Nutrition,* **7,** 481 (1934) (skin temperature).
28. Bohenkamp, H., *Verhandl. deut. Ges. inn. Med.,* **45,** 365 (1933).
29. Hardy, J. D., *J. Clin. Investigation,* **8,** 593 (1934) (surface radiometer).
30. Hardy, J. D., *J. Clin. Investigation,* **8,** 605 (1934).
31. Richards, C. H., and J. D. Hardy, *Federation Proc.,* **7,** 102 (1948) (an instrument for total thermal radiation).
32. Richards, C. H., S. Wolf, and H. G. Wolff, *J. Clin. Investigation,* **20,** 440 (1941).

33. Arrhenius, S., *Quantitative Laws in Biological Chemistry*. London, 1915.
34. Crozier, W. J., *J. Gen. Physiol.*, **7,** 189 (1924) (biological oxidations as a function of the temperature).
35. Crozier, W. J., *J. Gen. Physiol.*, **9,** 531 (1926).
36. Burton, A. C., *J. Cellular Comp. Physiol.*, **9,** 1 (1936) (principle of the master reaction); **14,** 327 (1939) (properties of the steady state).
37. Glasstone, S., K. J. Laidler, and H. Eyring, *The Theory of Rate Processes.* 1st ed., McGraw-Hill, New York, 1941.

CALORIMETRIC MEASUREMENTS

MAX KLEIBER, *University of California*

A. PRINCIPLES

1. Heat and Temperature

Heat and temperature were still confused in Newton's works. In 1750 Richman formulated a concept similar to what is now known as heat quantity, namely, mass times difference in temperature, although he used the term "calor" for this product as well as for temperature. Joseph Black (1728–1799) clarified the relation between temperature and heat, introducing a term "capacity for heat" for a concept now known as "specific heat," whereas "heat capacity" now designates the product of specific heat and mass of a body. The historical development of these concepts is given by Mach (1) and by Maxwell (2). Even today the terminology is still confused in some publications. "Thermogenic" is sometimes used instead of "calorigenic" to designate an action that increases the rate of heat production. A million calories of net energy in animal nutrition is called a "therm" (3). This is as unfortunate as the use of the same term "one therm" for the zero to one degree calorie (4). In some very well established expressions such as "thermochemistry" and "thermodynamics," "therm" to be sure still retains the double meaning of dealing with heat as well as temperature, but that is no reason why such confusion should be carried on in newly coined terms.

Temperature is discussed in another chapter of this book. Heat (calor) is the product of a difference in temperature and the heat capacity of a body.

2. Heat Capacity

Once a temperature scale is established, one may determine heat capacity by mixing substances with different temperatures and observing the resulting temperature of the mixture. The following discussion is simplified by the assumption that heat capacity does not change with changes of temperature.

If m grams of water at a temperature T_1 is mixed with m grams of water at a temperature T_2, the mixture will have the temperature of $(T_1 + T_2)/2$. The warm water lost as much temperature as the cold water gained. If, however, m grams of water at T_1 is mixed with twice as much water at T_2, the temperature of the mixture is closer to T_2 than to T_1. Generally one may formulate:

$$\frac{T_m - T_1}{T_2 - T_m} = \frac{m_2}{m_1} \quad \text{or} \quad (T_m - T_1)m_1 = (T_2 - T_m)m_2 \tag{1}$$

Temperature loss times mass of the warm portion is equal to temperature gain times mass of the cold portion.   The ability to raise the temperature of a given amount of cold water is proportional to the mass of the warmer water.

When different substances are mixed, the relation is more complicated.   If, for example, m grams of water at T_1 is mixed with m grams of steel at T_2 the rise in temperature of the water is only $\frac{1}{10}$ of the loss in temperature by the steel.   For a given difference in temperature, therefore, a unit mass of steel has only $\frac{1}{10}$ as much "ability" to raise the temperature of a given amount of a cold substance as does a unit mass of water.   To account for this different caloric behavior of different substances one uses the concept "specific heat," c.   The result of a mixing trial with two different substances at different temperatures may then be expressed as follows:

$$(T_m - T_1)m_1 c_1 = (T_2 - T_m)m_2 c_2 \tag{2}$$

where T_1, T_2 = original temperature of components, T_m = temperature of the mixture, m = mass (in grams), and c = specific heat.

For mixtures of several (i) components one may generally formulate:

$$T_m = \Sigma T_i m_i c_i / \Sigma m_i c_i \tag{3}$$

where the product of mass and specific heat $(m_i c_i)$ represents the heat capacity.   The final temperature of a mixture is thus equal to the sum of all the products of temperature times heat capacity of the components.   The latter sum is the heat capacity of the mixture.

To determine specific heats one may formulate from equation (2):

$$c_2 = c_1 \frac{m_1(T_m - T_1)}{m_2(T_2 - T_m)} \tag{4}$$

The specific heat of water is arbitrarily chosen as unity, thus if the substance 1 in equation (4) is water, $c_1 = 1$.   Heat capacities and specific heat are then expressed in terms of the heat capacity of 1 gram of water.

The simplifying assumption that heat capacity is independent of temperature is only approximately correct.   One might define a temperature scale, say between the freezing and the boiling points of

water, based on the postulate that the heat capacity of water within this range remains constant, but this temperature scale would deviate slightly from the one generally in use, namely, the thermodynamic scale. Based on the usual temperature scale the heat capacity of air-free water changes from 1.00738 cal. per gram at 0°C. to 1.00002 at 14°, to 0.99795 at 35°, to 1.00000 at 65°, and to 1.00697 at 100°C. (7).

Accurate data on specific heats of numerous substances are compiled in the *International Critical Tables* (5), the Smithsonian tables (6), and various other handbooks (7).

A number of rules permit the estimation of heat capacities (8). For monatomic ideal gases, the molar heat capacity is 3 cal. per degree at constant volume and 5 cal. per degree at constant pressure. For other gases the molar heat capacity increases with the number of atoms in the molecule. For solid elements the specific heat may be estimated according to the law of Dulong and Petit, which states that the product of specific heat and atomic weight of solid elements is 6.4.

For animals one may estimate:

heat capacity = grams water + 0.4 × grams dry matter   (5)

Disregarding ash content, one would estimate for an extremely lean animal with 25% protein and 75% water a specific heat of 0.85 cal. per gram. For a very fat animal with 30% fat, 20% protein, and 50% water, the specific heat would amount to 0.70 cal. per gram. A value of 0.8 cal. per gram appears to be a fair approximation for an average animal and was used by Rosenthal (9) in 1889. Burton (10) uses 0.83 as specific heat for human beings. Blood has a specific heat of 0.9 cal. per gram, which is close to the approximation represented by equation (5).

3. Heat Quantity

Temperature and heat capacity having been established, the quantity of heat (ΔQ) may be defined as the product of temperature change and heat capacity:

$$\Delta Q = (T_2 - T_1)mc \qquad (6)$$

where ΔQ = amount of heat transferred. This equation, however, is applicable only over a temperature range in which the specific heat

may be regarded as constant. When the specific heat changes with changes in temperature, the heat quantity may be expressed by the more general equation:

$$\Delta Q \ = \ m \ \int_{T_1}^{T_2} c_T \, dT \tag{7}$$

The specific heat (c_T) in this case is not a constant but is a function of the temperature.

Usually heat is measured in calories (cal.). In thermodynamics one uses as a rule the 15 degree calorie (11). This calorie is the amount of heat that raises the temperature of 1 gram of water from 15 to 16°C. It is almost exactly equal to the "mean calorie," namely, the amount of heat that raises the temperature of 0.01 gram of water from the freezing point to the boiling point at 1 atmosphere pressure.

Because electrical units can be standardized more easily than the heat capacity of water, the calorie is now defined on the basis of the international joule. 1 cal. = 4.1833 international joules. The joule in turn is defined as volt times coulomb. The volt is 1/1.0183 of the electromotive force of the normal Weston cell at 20°C., and the coulomb is the amount of electricity that deposits 0.001118 gram of silver in an electrolytic cell. This is 1/96494 of the electricity carried by 1 gram equivalent or the charge of 6.24×10^{18} electrons. Some authors express heat directly in terms of joules (5). It might be advantageous to maintain calories for expressing heat and use joules when other forms of energy, particularly chemical energy or radiant energy, are measured calorimetrically.

In much biophysical work heat quantities are given in kilocalories (kcal.). One kilocalorie equals one thousand calories. Terms such as "kilogram calories" and especially "large calories" are obsolete and should be avoided.

4. Latent Heat

The final temperature of a mixture of cold and warm substances may be the same as the initial temperature of one component. If, for example, 100 g. of water with a temperature of 8° are poured into 100 g. of water which contains a considerable amount of ice and has a temperature of 0°, the final temperature of the mixture is 0°. The 100 g. of the warmer water being cooled from 8 to 0° lost 800 cal. of heat. Instead of raising the temperature of the colder water this

TABLE I
Energy Equivalents (compare in lines)

Type of Energy	Unit	Absolute		Electrical		Mechanical		Heat
		Erg	Joule	Kw. hr.	Hp. hr.	Meter kg.	Liter atm.	Calorie[a]
Absolute.............	Erg	1	10^{-7}	$2.8 \cdot 10^{-14}$	$3.8 \cdot 10^{-14}$	$10.2 \cdot 10^{-9}$	$0.99 \cdot 10^{-9}$	$23.9 \cdot 10^{-9}$
	Joule	10^{7}	1	$0.28 \cdot 10^{-6}$	$0.38 \cdot 10^{-6}$	0.102	$9.9 \cdot 10^{-3}$	0.239
Electrical............	Kilowatt hour	$36 \cdot 10^{12}$	$3.6 \cdot 10^{6}$	1	1.36	$3.67 \cdot 10^{5}$	$35.5 \cdot 10^{3}$	$0.86 \cdot 10^{6}$
	Horsepower hour	$26.5 \cdot 10^{12}$	$2.65 \cdot 10^{6}$	0.736	1	$270 \cdot 10^{3}$	$26.2 \cdot 10^{3}$	$633 \cdot 10^{3}$
Mechanical...........	Meter kilograms	$98.1 \cdot 10^{6}$	9.81	$2.72 \cdot 10^{-6}$	$3.7 \cdot 10^{-6}$	1	0.097	2.34
	Liter atmosphere	$10.1 \cdot 10^{6}$	101.3	$28.2 \cdot 10^{-6}$	$38.3 \cdot 10^{-6}$	10.3	1	24.2
Heat.................	Calorie[a]	$41.8 \cdot 10^{6}$	4.18	$1.16 \cdot 10^{-6}$	$1.58 \cdot 10^{-6}$	0.427	$41.3 \cdot 10^{-3}$	1

[a] 1 calorie = heat that raises temperature of 1 gram of water from 14.5 to 15.5 °C.

heat melted 10 g. of ice. This heat is, so to speak, stored in the mixture; it can become manifest again when, for example, 100 g. of steel at an initial temperature of $-80°$ is added to the mixture. Then 10 g. of water will freeze and raise the temperature of the steel from -80 to $0°$. Between being lost by the warmer water and finally gained by the cold steel the 800 cal. of heat have passed through a hidden state, as it were. This state is termed "latent." One distinguishes thus between "sensible heat," which is observable as a temperature change, and "latent heat," which produces no change in temperature but instead is involved in changes of state—solid to liquid or liquid to gas.

The latent heat is used extensively in calorimetry. Famous examples are the ice calorimeters of Lavoisier-Laplace and of Bunsen.

5. Chemical Energy

The latent heat of change of state is mainly a matter of the spatial relations between molecules. A further form of latent heat is involved in chemical changes, that is, changes in the arrangements of atoms in the molecules. When 1 g. of glucose is oxidized to 1.47 g. of carbon dioxide gas and 0.60 g. of liquid water, 3.74 kcal. of heat is produced. This is the "heat of reaction," or in this particular case the "heat of combustion." For given initial and final states of the process a definite amount of heat is developed from the oxidation of a given amount of glucose. This amount remains the same whether the process takes place as an explosion in a calorimetric bomb or as a series of enzymic processes such as occur in animal tissues. The generalization of this example is known as the fundamental law of thermochemistry, the law of constant heat sums, or the law of Hess, who announced it in 1840.

Three years later Robert Mayer formulated an even broader generalization including the relation of heat and work. As, in some mixing trials, sensible heat, which apparently disappeared, may be conserved as latent heat of state, so can mechanical work of motion be conserved as potential work of position, for example in a swinging pendulum. When, however, the pendulum is stopped at its lowest position, its work of motion as well as its potential work of position have actually been lost. This loss of work is accompanied by the production of heat. Work of motion can thus become latent not only as work of position but also as heat; conversely heat can become

latent not only as latent heat of melting or as heat absorbed in a chemical change but also as mechanical work.

Some agent remains constant in all these changes. This agent is called energy. Work, light, electric potential, and, since Einstein, also mass are various manifestations of this agent. Heat can be produced from work or electric energy from heat, but energy itself can neither be produced nor destroyed; it can only change the form of its manifestations. Expressions such as "energy production," rather frequently used by biochemists, are therefore unsound. The law of conservation of energy includes, as a special case, the fundamental law of thermochemistry mentioned above.

On the basis of the law of conservation of energy, or the first law of thermodynamics, any form of energy may be measured in terms of any other. Table I gives the conversion factors for various units of energy.

B. FIELDS OF APPLICATION

For biophysics, calorimetry is applied in research on heat transfer of organisms. Of equal importance is the application of calorimetry for the measurement of other forms of energy. Transformations to any form of energy but heat are as a rule incomplete. The transfer of any form of energy to heat, on the other hand, can ordinarily be performed easily and completely. This unique position of heat among the various manifestations of energy makes calorimetry particularly suitable for the measurement of other forms of energy. Of special interest for biophysicists are the calorimetric measurements of radiant energy and of chemical energy.

1. Heat Transfer

The knowledge of animal heat transfer is important for the choice of standard clothing. It is widely applied also in the design of insulation and of heating and cooling devices in houses for man and animals. Builders of hatcheries, brooders, and poultry houses are particularly interested in the rates of heat transfer of the birds. Partitional calorimetry—that is, the measurements of radiation, convection, conduction, and water evaporation as parts of total heat transfer—proves especially helpful. The knowledge of the thermostatic characteristics of farm animals will be of importance not only for the care and management of the herds, but also in the selection for breed-

ing of animals best fit for given climatic conditions. Attempts are underway to combine the high rate of production and efficiency of food utilization for meat and milk inherent in our modern western breeds of cattle with the great ability to endure high environmental temperatures, which characterizes Indian breeds (Brahman cattle). A standard scale for measuring thermostatic ability would undoubtedly be advantageous for this breeding work.

2. Measurement of Radiant Energy

Radiant energy is often measured by transforming it into heat (*12*). The transformation takes place as a rule in a so-called black body, and the heating effect is measured either by rotation of a paddle wheel, as in Crooke's radiometer, by the change in temperature of the arm of a Wheatstone bridge, as in Langley's bolometer, or as a thermoelectric potential of a thermocouple, as in Boys' radiomicrometer or in Hardy's radiometer for measuring radiation rates of human skin (*13*). Measurements of radiant energy are basic in research on photosynthesis (*14*). They are also important in the study of heat exchange of animals and man.

3. Measurement of Chemical Energy

The most widespread application of calorimetry in biological research is the measurement of chemical energy. Among the innumerable possible aspects of life, the one that shows life as a transformation of energy has proved to be particularly fruitful. In the great process: radiant energy $\rightarrow$ chemical energy $\rightarrow$ heat, the first step is of particular importance in plant physiology, the second in animal physiology.

From the sun's great stream of radiant energy that flows down to the level of heat, organisms channelize a streamlet and retard the degradation process to heat, using this radiant energy for the synthesis of organic compounds. These compounds in turn furnish energy necessary for the life of the allotrophic organisms, particularly animals.

The efficiency of plant production may be expressed in terms of calories of chemical energy yield per acre or per calorie of radiant energy received on the ground. Food requirements and performances of work of man and animals are measured in calories, and important deductions in biochemistry are based on free energy (*11,36*).

TABLE II.   Classification of Calorimeters[a]

Category of classification	Type of calorimeters					
	No heat transfer to outside, adiabatic		Heat transfer to outside			
Heat						
Temperature of calorimeter	Constant isothermal	Changing temperature	Temperature in calorimeter nearly constant			
Major form of heat transferred	Latent heat	Sensible heat	Heat flow through walls		Heat flow in circulating medium	
Quantity measured	Amount of substance melted or evaporated	Difference in temperature of calorimeter in time	Difference in temperature inside and outside of calorimeter as index for heat flow	Heat produced in compensating chamber so as to match thermal head of measuring chamber	Difference in temperature at entrance and exit of circulating medium and rate of flow of this medium	
Calculation of results	km	$C \Delta T_1$	$L \Delta T_2$	$i^2 r$	$c \dfrac{\Delta m}{\Delta t} \Delta T_3$	
Examples	Ice calorimeter	Bomb calorimeter, Hill's microcalorimeter	Richets and Rubner's respiration calorimeter	Compensation calorimeter	Cooling medium: Circulating water — Respiration calorimeter of Atwater and Rosa, Armsby	Circulating air — Respiration calorimeter of Murlin and Burton, of Auguet and Lefèvre

[a] k = latent heat per unit mass of melted or evaporated substance. m = mass of substance melted or evaporated. C = heat capacity of calorimeter (without jacket). ΔT_1 = increase in temperature of calorimeter (without jacket). L = heat transfer coefficient of calorimeter wall. ΔT_2 = difference in temperature inside and outside of calorimeter wall. i = intensity of electric current in heater of compensating chamber. r = electrical resistance of heater in compensating chamber. c = specific heat of circulating medium. $\Delta m / \Delta t$ = rate of flow of circulating medium. ΔT_3 = increase in temperature of medium by circulating through chamber.

C. CALORIMETRIC METHODS

An account of the development of calorimetric methods is found in Saha's treatise (15). A thorough discussion of the present status of the subject is found in a chapter by Sturtevant (16).

1. Classification and Choice of Calorimeters

Table II indicates how calorimeters may be classified according to their major characteristics.

The choice of a calorimetric method may have to be based on criteria different from those used in Table II. The magnitude of thermal effects may have to be considered. The measurement of a man's rate of heat production requires macrocalorimetry. The study of the heat of setting concrete even in giant dams, on the other hand, may well be accomplished by microcalorimetric methods because the essential conditions such as temperature, pressure, and chemical composition may be adequately reproducible on a small scale. Availability and cost of equipment, skill, and time necessary for the measurement may be other factors involved in the choice of the method.

Heats of combustion of ordinary plant or animal materials may be accomplished by rather well standardized methods and a number of bomb calorimeters for this purpose are readily available and adequately described in the catalogs of the supply houses for laboratory equipment. Even without previous training in physics or chemistry, normally intelligent people can be taught in a few weeks to operate these instruments for satisfactory routine measurements, provided the teacher and supervisor of these technicians has himself sufficient basic knowledge. For such routine measurements by technicians with limited training, a calorimeter jacket with high heat capacity and nearly constant temperature (see Sect. C7) is preferable to the adjustable adiabatic jacket. The three rows of temperature figures obtained with the constant jacket in the preliminary, main, and after period of the measurement can readily be checked, and irregularities are easily detected, whereas errors in the operation of the adiabatic jacket may be difficult to trace after the measurement is completed (see the last paragraph in Section C7).

For the measurement of heat exchange during photosynthesis or for animal calorimetry no generally followed routine is established and no standard equipment is available. Good training in physics and rather well developed engineering ability are among the prerequi-

sites for leaders of research in these fields, because they have to design apparatus and develop methods for their particular problems. Calorimetric frontier work such as myothermic and neurothermic measurements in periods measured in milliseconds taxes to capacity a well trained and critical mind of a physicist combined with the ingenuity of a clever engineer.

2. Adiabatic Measurements

Adiabatic measurements are those in which there is no energy exchange with the environment. Typical representatives of adiabatic apparatus are the ice calorimeters. The heat produced or released in the instrument is measured by the amount of ice melted. The result may be calculated as follows:

$$\Delta Q = 79.71 M \tag{8}$$

where ΔQ is the heat developed (in calories) and M is the mass of ice melted at $0°C$. (in grams).

Lavoisier and Laplace measured the amount of ice melted by weighing the water that ran off from a pack of ice surrounding the source of heat, such as a guinea pig. The ice pack was surrounded by a jacket of water and ice.

The amount of ice melted may be measured more accurately by a decrease in volume. The relation between the decrease in volume and the amount of ice melted may be stated as follows:

$$M = \Delta V \Big/ \left(1.000 - \frac{1}{0.917}\right) = 11.05 \; \Delta V \tag{9}$$

where M symbolizes the amount of water melted in grams, ΔV the decrease in volume of the ice water mixture, 1.000 the density of water at the freezing point, and 0.917 the density of ice at the freezing point. Substituting this relation for M in equation (8) one finds finally:

$$\Delta Q = 79.71 \times 11.05 \Delta V = 880.8 \Delta V \tag{10}$$

The use of the contraction in volume by melting of ice as a calorimeter is credited to Hermann, a Russian scientist, who invented this method in 1834. In 1870 Bunsen applied this principle for his well known calorimeter, which is used mainly for measuring specific heat. For a thorough discussion of the precautions necessary in the use of

Bunsen's ice calorimeter, and particularly the use of this apparatus for microcalorimetry, the reader is referred to the recent book by Swietoslawski (17).

Combustion calorimetry by the use of the calorimetric bomb may be classified among adiabatic measurements, even if no adiabatic jacket is used, because the heat flow between calorimeter proper and environment (or jacket) plays only a minor role.   It is then designated as "heat leak" and, in the final calculation, figures as a correction term.   Heat leak and combustion calorimetry will be discussed in the following sections.

3. Measurement of Heat Flow through Walls

Instead of preventing heat flow as in adiabatic calorimeters, one may measure it on the basis of Fourier's law, which, for steady state, may be formulated as follows:

$$\Delta Q/\Delta t = \lambda S(T_i - T_s)/d \tag{11}$$

where $\Delta Q/\Delta t$ = rate of heat flow (for example, calories per minute), λ = heat conductivity of insulating layer, d = thickness of insulating layer, S = area of heat stream—in this case, area of insulating surface layer, T_i = temperature of inner surface of insulating layer, and T_s = temperature of outer surface of insulating layer.

By empirical calibration one may determine a heat conduction constant, $L = \lambda S/d$, for any given calorimeter, and then formulate:

$$\Delta Q/\Delta t = L(T_i - T_s) \tag{12}$$

This principle of measuring heat flow was applied especially in early animal calorimetry by Richet and by Rubner (18,19).   In their calorimeters the animal room is surrounded by a layer of air (A_i, Fig. 1). This air space is surrounded by an insulating layer I, which in turn is enclosed in an outer layer of air, A_s.   Each of the two air layers is hermetically sealed and is connected to the other through a manometer, M.   After an empirical calibration the difference in pressure, p, between the two air layers indicates the difference in temperature, $T_i - T_s$, for the calculation of heat flow.

Instead of using an air thermometer for measuring $T_i - T_s$, one might nowadays advantageously measure this difference with a set of thermoelectric junctions distributed on the inner and outer walls of the insulating layer.
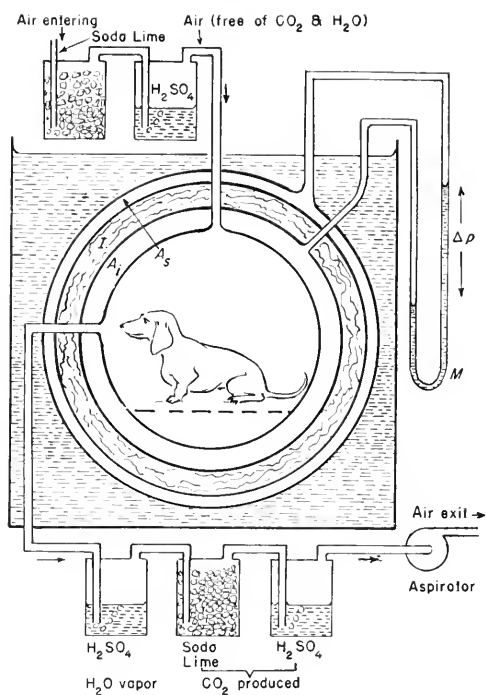
Fig. 1.   Diagrammatic illustration of respiration calorimeter of Rubner.
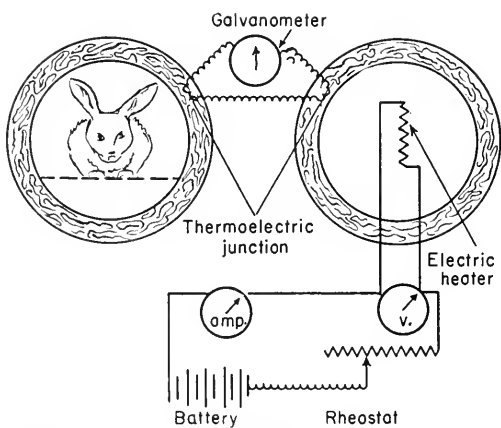
Fig. 2.   Schematic diagram of an electric compensation calorimeter.

4. Differential Calorimetry

Instead of measuring the heat flow from a calorimeter chamber as described above, one may match this flow by an equal flow in a quasi-equal second chamber, the "compensating chamber." The rate of heat production in the calorimeter chamber is then measured by the rate of heat production in the compensating chamber necessary to produce an equal "thermal head," $T_i - T_s$. The arrangement of a differential calorimeter, also called a compensating calorimeter, is illustrated in Figure 2.

Since, by definition of "quasi-equal conditions" for both chambers, T_s is the same for both, it is only necessary to match the two inside temperatures. A galvanometer is connected to thermoelectric junctions distributed on the inside walls of both chambers. It serves as a null instrument, whose sensitivity may be very high. The heat flow in the compensating chamber is usually produced by an electric current whose intensity and potential drops may be measured very accurately. The rate of heat production in the calorimeter chamber is then calculated as:

$$\Delta Q / \Delta t = 0.239 i E \qquad (13)$$

where $\Delta Q / \Delta t$ is the rate of heat production in calories per second, i is the electric current in amperes, and E is the potential difference in volts.

Detailed discussions of differential calorimetry are found in the literature (20,21). For experiments on a large scale a major source of error appears to be the difficulty in making the thermal conditions in the two chambers equal. If this equality for two chambers, used simultaneously, is not more accurate than the equality of the thermal conditions of the same chamber used at different periods, ordinary calorimetry may be as accurate as differential calorimetry. For an appraisal of the differential method in microcalorimetry the reader should consult the recent book by Swietoslawski (17).

5. Prevention of Heat Leaks

Instead of measuring the heat flow through the walls of the calorimeter, one may prevent this flow or decrease it to a small fraction of the measured heat, which is then called a "heat leakage."

For experiments of long duration and relatively steady rate of heat production, heat leakage may be reduced to an insignificant rate

by good insulation alone (*i.e.*, make L in equation 12 small).   This was accomplished in Capstick's calorimeter for large animals (*22*).

Decrease or prevention of heat flow through walls may also be accomplished by adiabatic jackets.   An adiabatic jacket consists of two concentric metal walls with air or a liquid between them.   When the two walls are kept at the same temperature, $T_i - T_s$ in equation (11) or (12) is zero and no heat flows.   This condition is accomplished by a series of thermoelectric junctions distributed over inner and outer
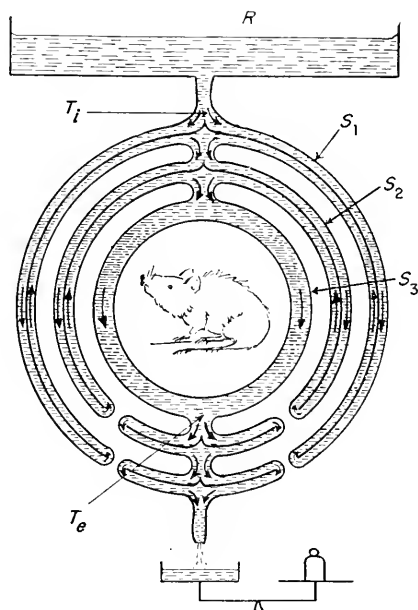
Fig. 3.   Schematic diagram of "labyrinth" flow
calorimeter of Swietoslawski.

wall, and facilities to heat or cool the outer wall so that the galvanometer connected with the thermopiles shows no potential difference. This system was invented by Rosa and applied in the famous respiration calorimeter for human beings of Atwater and Rosa (*23*) (Fig. 4) and that for cattle built by Armsby and Fries (*3*).

Prevention of heat leaks is particularly important for microcalorimetry, which would hardly be possible without the very good insulation provided by Dewar vessels.   Adiabatic jackets are also used in microcalorimetry.   A most ingenious device for reducing heat leak-

age is the labyrinth flow jacket, developed mainly by Swietoslawski (17). Figure 3 schematically illustrates the principle of a labyrinth flow calorimeter.

Water at the environmental temperature T_i flows from the reservoir R down through the outer shell S_1, returning in a counterflow through the inner layer of shell S_1 to the top. In a similar way the water passes the middle shell S_2 and then enters the inner shell S_3. In the inner shell the water takes up the heat given off by the animal. The difference in temperature of the water entering the calorimeter (T_i) and that leaving the inner shell (T_e) multiplied by the rate of flow is the rate of animal heat loss. Some correction for changes in moisture content and temperature of the air may be necessary.

The labyrinth flow effectively reduces heat leakage since it keeps the temperature of two adjoining walls nearly equal. Most of the heat that might flow by conduction within the material of the shells from inside to outside would be picked up by the inflowing water and thus returned to the interior. The reverse would be true for the exit manifold whose material therefore should have a low heat conductivity.

6. Measurement of Heat Flow by Circulating Medium

The prevention of heat leaks through the walls does not necessarily make a calorimeter adiabatic. The claim of Auguet and Lefèvre (24) that their calorimeter for human beings was both perfectly adiabatic and isothermal is an error. Heat produced in their calorimeter is carried outside by a cooling stream of air. The same method is used in a more recent human calorimeter by Murlin and Burton (25). In the classical respiration calorimeter of Atwater and Rosa (23) the cooling medium is circulating water. Figure 4 illustrates the principle of this apparatus.

The subject is surrounded by two concentric walls. The outer wall is always kept at the same temperature, T_s, as the inner wall so that there is no heat leakage. The heat produced by the subject is carried away by the water that flows into the chamber at the temperature T_{wi} and leaves the chamber with a temperature T_{we}. The rate of flow is measured by the balance B.

The rate of heat loss in this case may be calculated as follows:

$$\frac{\Delta Q}{\Delta t} = s \frac{\Delta W}{\Delta t} (T_{we} - T_{wi}) \tag{14}$$

where $\Delta Q / \Delta t$ = rate of heat loss in calories per second, s = specific heat of medium, water in this case, in calories per gram per degree Centigrade, $\Delta W / \Delta t$ = rate of flow of the medium, in grams per second, T_{we} = temperature of outflowing water, and T_{wi} = temperature of inflowing water.

If the weight of the water, W, is measured in metric units, s is nearly unity (depending slightly on the temperature).
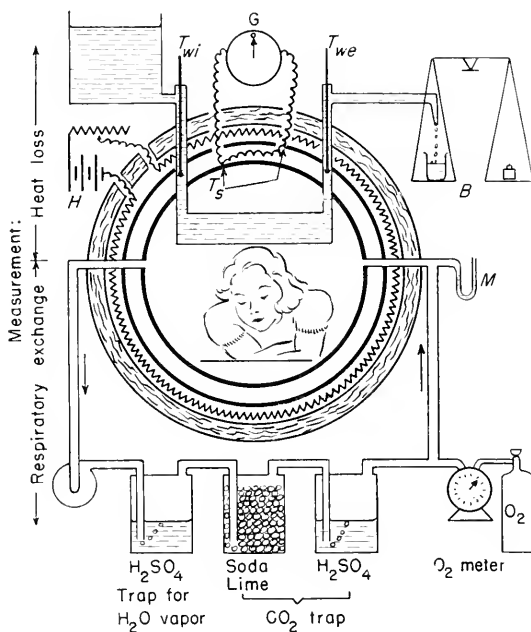
Fig. 4.   Schematic diagram showing principle of Atwater-Rosa, respiration calorimeter.

To the heat carried away by the water must be added the heat given off as latent heat of water vaporization.   For a temperature of $20°$ this latent heat is:

$$Q_c = 585 \, W_e \tag{15}$$

where Q_c = latent heat of evaporation in calories and W_e = grams of water evaporated.

The latent heat of evaporation for other temperatures is found in the literature (5–7).

The amount of evaporated water is measured by the increase in

weight of the bottle containing concentrated sulfuric acid through which the air from the chamber bubbles, as indicated in Figure 4. Under normal conditions the latent heat of water evaporation amounts to about one-fifth of the total heat given off by man.   Its share of the total heat loss increases with increasing environmental temperature, and when the environment is at body temperature all the heat produced has to be given off as heat of evaporation.   The air-circulating system of the Atwater-Rosa apparatus permits the measurement of carbon dioxide production and of oxygen consumption as well as water evaporation, as indicated in Figure 4.

7. Bomb Calorimetry

The energy content of substances used as fuel for furnaces or organisms is measured by the heat evolved when these substances are combusted.   Aside from measuring fuel values, however, heats of combustion serve as the most common bases for calculating other important data for thermochemistry and thermodynamics, such as heat of reaction and free energy.   Combustion calorimetry is therefore the type of heat measurement most widely used in bioenergetics. Among combustion calorimeters, in turn, the instrument best known and most widely used is the bomb calorimeter, invented by Berthelot in 1869.

Figure 5 is a schematic sketch of a simple bomb calorimeter.   The sample, F, to be combusted is weighed into a little dish, which is subsequently suspended in the center of the bomb, B.   For ignition, one end of a thin iron wire, I, is clamped to an electrically insulated plug in the wall of the bomb; the other end of the wire is directly connected to the wall of the bomb.   In many laboratories a weighed cotton thread is hung over the ignition wire for more reliable ignition of the sample.

The bomb is then closed—as a rule by a heavy steel cap screwed on a lead gasket, and filled with oxygen to a pressure of 30 atmospheres through valve V.   The bomb is then placed into the calorimeter pail, which contains a weighed amount, usually 2 to 3 kg., of water, W.   This pail, in turn, is installed inside the calorimeter jacket, J, which is filled with water to provide a large heat capacity and consequent uniformity of the temperature surrounding the calorimeter water, W.   The jacket may be covered with an insulating layer to shield the inside from sudden changes in room temperature. This precaution is unnecessary when calorimetric measurements are carried out in a temperature-controlled room.   A good stirrer, S, is essential.   A cover, C, to prevent errors from water evaporation is very desirable.   The temperature of the calorimeter water is measured by a thermometer, T,

which may be of the Beckmann type, or may be an absolute thermometer graduated to 0.01 °C. in order to permit the estimation of thousandths of a degree. After a steady thermal state is reached during a preliminary period, a small electric current is passed through the ignition wire, heating it so that it burns and ignites the sample.

Temperature leaks during the experiment may be estimated from temperature readings at regular intervals in a preliminary period and a period of steady change after the combustion. As a rule, five minutes will be sufficient time for each of the three periods: preliminary period prior to ignition, main period just after ignition, and after period, following the main period.
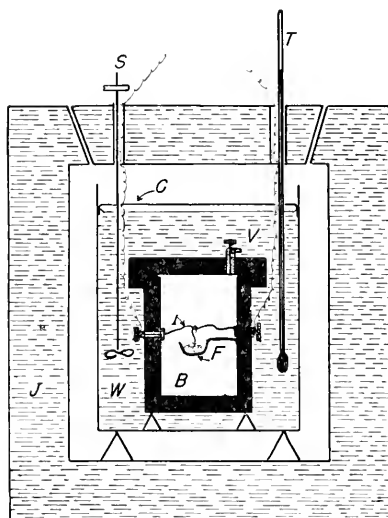
Fig. 5. Bomb calorimeter in cross section.

A correction for temperature loss may be calculated on the basis of Newton's law of cooling, which indicates the following relation:

$$V = K(T - T_s) \qquad (16)$$

where V = rate of temperature loss in degrees Centigrade per minute and K is the cooling coefficient or cooling rate per degree difference. The difference $T - T_s$ is known as "thermal head," where T is the temperature of the calorimeter water and T_s is the temperature of the environment (jacket).

Application of this law leads to the following correction equation:

$$\sum_{0}^{n} V = n\bar{V}_a + \frac{\bar{V}_z - \bar{V}_a}{\bar{T}_z - \bar{T}_a}\left(\frac{T_0 - T_n}{2} + \sum_{1}^{n-1} T - n\bar{T}_a\right) \quad (17)$$

where $\sum_{0}^{n} V$ = total temperature loss in main period = temperature correction, n = number of intervals in main period (as a rule, number of minutes), $\bar{V}_a$ = mean rate of temperature loss in preliminary period (degrees Centigrade per minute), $\bar{V}_z$ = mean rate of temperature loss in after period, $\bar{T}_a$ = mean temperature of calorimeter water during preliminary period, $\bar{T}_z$ = mean temperature of calorimeter water during after period, T_0 = temperature of calorimeter water at ignition time, T_n = last temperature reading of main period, and T = temperature of calorimeter during main period. This is the calorimeter correction formula of Regnault-Pfaundler. For its derivation one has to assume that the environmental temperature T_s, which in a well constructed calorimeter is equal to the temperature of the jacket, remains constant. This constancy may be approximated sufficiently by using jackets of sufficient heat capacity. The requirement may be estimated as follows:

One may assume for an ordinary bomb calorimeter a cooling coefficient of 0.002° per minute per degree Centigrade (unpublished result by author). For a duration of five minutes and an increase in temperature of the calorimeter of 5° the temperature loss would amount to 0.05°C. Assuming a heat capacity of the calorimeter of 2.5 kcal. per degree Centigrade, the temperature loss represents a heat loss of $0.05 \times 2500 = 125$ cal. If this amount of heat were to raise the temperature of the jacket only 0.001°, the jacket should have a heat capacity of 125 kcal. (125/0.001) or an equivalent of 125 liters of water. If a change of 0.005°, which amounts to 0.1% of the result, is allowable, a jacket of 25 liters capacity will do.

Errors from inaccurate assumptions, such as the constancy of environmental temperature, may be eliminated to a considerable extent if the calorimeter is calibrated by combusting a reference substance under conditions similar to those prevailing during subsequent measurements. Benzoic acid is supplied for this purpose by the National Bureau of Standards. Salicylic acid is used in the Swiss Institute for Fuel Investigations.

A more elaborate discussion of the Regnault-Pfaundler formula and its improvements is found in White's monograph (33). For most determinations of heats of combustion by bomb calorimetry the old formula is sufficiently accurate.

To avoid corrections for heat leakage some bomb calorimeters are supplied with so-called adiabatic jackets, that is, jackets that may be heated in an attempt to make their temperature equal to the temperature of the calorimeter water. Such jackets are especially suitable for measuring uniform continuous thermal processes such as the heat evolved by a radium preparation, or during the hardening of Portland cement. For the abrupt heat production of explosions, however, such as those in the calorimeter bomb, the advantage of an adiabatic jacket is doubtful. Even with adequate stirring of the jacket water and accurate measurement of its temperature, it is difficult to make the temperature of the inner jacket surface always equal to that of the outer surface of the calorimeter bucket. Some of the commercially obtainable bomb calorimeters have adiabatic jackets even without stirrer. The loss in reliability introduced in bomb calorimetry with such devices seems a rather high price for avoiding some simple calculation of temperature loss.

8. Microcalorimetry

Microcalorimetry deals with small thermal effects. In contrast to microchemistry, relatively large amounts of material may be involved in these measurements. Microcalorimeters may therefore be rather

TABLE III
Characteristics of Microcalorimeters

Characteristics	Inventor			
	Ward	Hill	Lange	Whipp (26)
Water equivalent of heat system, g.	50	0.6	1000	1
Number of thermocouples of main thermometer	31	150	1100	5
Temperature sensitivity, °C./mm. galvanometer deflection	4×10^{-6}	10^{-6}	0.6×10^{-6}	3×10^{-6}
Heat sensitivity, cal./mm. galvanometer deflection	2×10^{-3}	0.5×10^{-6}	0.6×10^{-3}	3×10^{-6}
Resistance of galvanometer, ohms	8.5	25	25	1
Resistance of thermocouples, ohms	11.5	50	75	1
Sensitivity of galvanometer, amp./mm. deflection	3.5×10^{-9}	3.3×10^{-10}	10^{-9}	3×10^{-10}
Insulation of couples	Organic	Organic	Organic	Glass

large; thus the twin calorimeter of Gucker for the measurement of
small heats of dilution in dilute electrolyte solutions had a volume of
1 liter for each of the two chambers. The table of Whipp (26) best
illustrates typical characteristics of some microcalorimeters (see Table
III).

Common to all these instruments is the use of electric thermo-
couples with a very high temperature sensitivity of about a micro-
degree per millimeter deflection of the galvanometer. The greatest
heat sensitivity is reached in Hill's instrument used for measuring heat
production of nerves in milliseconds. A deflection of 1 mm. of the
galvanometer spot indicates 0.5 microcalorie, a microcalorie being
one-millionth of a calorie. This heat sensitivity is about ten times
as high as the corresponding sensitivity in the measurement of oxygen
consumption reached in the Cartesian diver microrespiration appara-
tus, namely, one-millionth of a cubic milliliter of oxygen per milli-
meter of manometer change.

This high sensitivity of microcalorimeters makes processes im-
portant that have no noticeable effect on the result in macrocalorim-
etry. Thus Hill attributed large prolonged galvanometer deflec-
tions in his apparatus to "rainstorms" set off in the tiny chambers by
electric charges on nerve-stimulating wires. He also calculated
that a pinch of a rubber tube causing a volume change of one part
in a million would produce an adiabatic temperature change of 10
microdegrees, which in turn would make the galvanometer spot travel
50 cm. (27, p. 119).

Microcalorimeters require special insulation. Dewar vessels are
often used for this purpose. Very accurate regulation of environ-
mental temperature is especially important. Tian (28) has invented
a thermal damping system in which several concentric layers of al-
ternately good and bad heat conductors form a multiple jacket. The
temperature inside this jacket is nearly unaffected by temperature
fluctuations outside, so that the inside temperature represents very
accurately the mean environmental temperature. Trends of envi-
ronmental temperature penetrate the damping system very slowly.
With a Tian damper, Hill was able to decrease the "zero creep" of a
thermostat to about 2 microdegrees in ten minutes.

For detailed information and critical discussion of microcalorim-
etry the recent book of Swietoslawski should be consulted (17).
An earlier article by Meyerhof appeared in Abderhalden's handbook
(29).

9. Partitional Calorimetry and Kata Thermometer

For characterizing the thermostatic behavior of animals a partition of the total heat flow into its components—radiation, convection, and evaporation of water—is important. Rubner has already worked intensively on this problem and has found that for average comfortable conditions evaporation accounted for about 20% of the total heat loss, convection and conduction for 30%, and radiation for 50%. Such measurements have been extended and refined, especially by Hardy and Dubois and by a group of biophysicists at Yale University (30).

Hardy (13) has designed a radiometer that measures radiant energy after its conversion to heat at a black surface. One thermoelectric junction is installed on the blackened receiver of radiation and the other junction of the electric thermocouple is at a point shielded from the radiation. The difference in temperature, indicated by the electric potential between the two junctions, measures the intensity of radiation. On the basis of Stefan-Boltzmann's law of radiation, the radiometer may be used also to measure surface temperatures of animals.

The Hardy radiometer measures the surface temperature of animals in comparison with the temperature of a reference body. If this reference body is at the temperature of the entire environment such as the walls, floor, and ceiling of a room, the heat radiated from a square centimeter of the animal surface may be calculated on the basis of the Stefan-Boltzmann law:

$$\Delta Q_R/A\,\Delta t = \sigma(T_S^4 - T_R^4) \tag{18}$$

where $\Delta Q_R/A\,\Delta t$ = rate of heat loss by radiation per square centimeter of surface area, σ = Stefan-Boltzmann constant = 1.37×10^{-12} cal. per second per square centimeter, T_s = temperature of radiating surface, in degrees absolute, and T_R = temperature of receiving surface, in degrees absolute.

In the John Pierce Laboratory at Yale, radiation may be controlled by surrounding the experimental person with copper sheets that reflect radiant heat received from outside. The temperature of the walls in radiation exchange with the body may thus be varied independently of the temperature of the air surrounding the body.

The reverse tendency to the differentiation of the total heat loss of animals into the various paths is the integration of the thermal environmental factors into one figure. This integration is accom-

plished, to a certain extent, by the so-called Kata thermometer designed to measure the cooling power of various environments. A short historical account of this instrument, with the description of a new form, the thermointegrator, is given by Winslow and Greenburg (*31*).

A simple form of Kata thermometer is shown in Figure 6A. It is a thermometer with two temperatures, t_1 and t_2, marked on its stem.
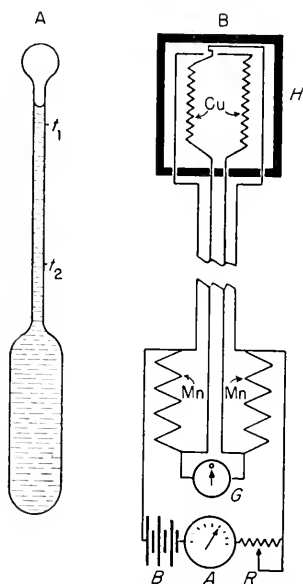
Fig. 6.   Schematic sectional diagrams of two
types of Kata thermometer.

The instrument is first heated above its upper temperature and the time required to cool it from t_1 to t_2 is measured with a stop watch. The cooling power of the environment is an inverse function of that cooling time.

Another type of Kata thermometer measures the heat necessary to maintain the temperature of a test body at a given level. Figure 6B shows an instrument of this type. The test body, H, contains two spirals of enameled copper wire, Cu, which are connected to two spirals of manganin wire, Mn, to form a Wheatstone bridge supplied by an electric current from the battery, B. The resistances are adjusted so that the bridge is in equilibrium, indicated by a null

reading on the galvanometer, G, when H is at a given temperature, say 37°C. Since the resistance of the copper wire varies with the temperature, and the resistance of the manganin wire is practically independent of temperature changes, the bridge is balanced only when H is at the temperature for which the instrument is adjusted. Since the resistance of the bridge in equilibrium is a constant, the heat necessary to keep the test body at the desired temperature may be calculated from the intensity of the current regulated by the rheostat, R, and read on the ammeter, A.

10. Indirect Calorimetry

Animal calorimetry measures the heat loss of animals. When the heat capacity and temperature of a body remain constant, the rate of heat loss is equal to the rate of heat production. In short experiments heat production and heat loss may differ significantly.

Animal heat production is now usually estimated from chemical measurements. This procedure is known as indirect calorimetry. It is based on the law of Hess, mentioned above, and deals only with initial and final conditions, being unaffected by the particular processes known as intermediary metabolism, which characterize the change. A given decrease of carbohydrate, protein, and fat in the animal body and a given consumption of oxygen, accompanied by a corresponding production of excretory products, such as urea and carbon dioxide, indicate a definite heat production.

The heat of combustion of the major group of nutrients is given in Table IV. The heat of combustion is, as a rule, measured in a

TABLE IV
Caloric Equivalents of Nutrients

Substance	Heat of combustion, kcal./g.	Catabolizable energy (physiological fuel equivalents), kcal./g.
Fat	9.5	9.5
Protein	5.7	4.8[a]
Carbohydrates	4.0	4.0

[a] Calculation for protein:

Heat of combustion per gram nitrogen in protein 5.7×6.25	$= 35.6$ kcal.
Loss in urine: heat of combustion per mole urea $= 152$ kcal.	
per gram nitrogen in urea $= 152/28$	$= 5.4$ kcal.
Difference: Catabolizable energy per gram nitrogen in protein	$= 30.2$ kcal.
Catabolizable energy per gram protein $= 30.2/6.25$	$= 4.8$ kcal.

bomb at constant volume (ΔE). The oxidation in animals, however, takes place at constant pressure. To obtain the heat of combustion at constant pressure, or enthalpy, ΔH, one has to subtract from ΔE the heat equivalent of the work of expansion against the atmosphere. For carbohydrates this work is zero since the volume of carbon dioxide produced is equal to the volume of oxygen consumed. Thus, in this case the enthalpy is equal to the heat of combustion measured in the bomb. In the combustion of fats, a greater volume of oxygen is consumed than of carbon dioxide produced. Under constant pressure there is, thus, a decrease in volume, and the heat equivalent of the work done by the atmosphere in compressing the system has to be added to the heat of combustion as measured at constant volume. The heat equivalent of this work, however, amounts to only 0.14% of the heat of combustion.

For proteins, the heat production in the calorimeter bomb is considerably greater than the heat the animal can derive from protein catabolism. A certain amount of the chemical energy of the protein is lost as chemical energy in urine. This loss is calculated and the resulting catabolizable energy equivalents are given in Table IV. Catabolizable energy or metabolizable energy measures the physiological fuel equivalents of nutrients.

TABLE V

Estimation of Heat Production from Fat and Protein Catabolism

R.Q.:	Heat production (in kcal.) per		
$\dfrac{\text{moles } CO_2 \text{ produced}}{\text{moles } O_2 \text{ consumed}}$	Mole CO_2 produced $= \dfrac{231.7}{R.Q.} - 179.6$	Mole O_2 consumed $= 231.7 - 179.6\ R.Q.$	Liter O_2 consumed
0.707	148	105	4.7
0.72	142	102	4.6
0.74	133	99	4.4
0.76	125	95	4.2
0.78	117	92	4.1
0.80	110	88	3.9

From the "respiratory quotient" (R.Q.) one may estimate the proportion of fat and protein that a starving animal metabolizes provided its glycogen content does not change significantly.   This is the rule in "postabsorptive condition."   From these results one may further calculate the caloric equivalents of carbon dioxide and oxygen for a given R.Q.   This is done in Table V.

If the nitrogen balance is measured, one may subtract from the

amount of carbon dioxide produced and oxygen consumed the part
related to protein metabolism and from the so-called "nitrogen-free
R.Q." of the remaining carbon dioxide and oxygen calculate how much

TABLE VI

Estimation of Heat Production from Fat and Carbohydrate Catabolism

R.Q.: $\dfrac{\text{moles } CO_2 \text{ produced}}{\text{moles } O_2 \text{ consumed}}$	Heat production (in kcal.) per		
	Mole CO_2 produced $= 113 + 84.45\left(\dfrac{1}{R.Q.} - 1\right)$	Mole O_2 consumed $= 84.45 + 28.55$ R.Q.	Liter O_2 consumed
0.707	148	105	4.7
0.75	141	106	4.7
0.80	134	107	4.8
0.85	128	109	4.8
0.90	122	110	4.9
0.95	118	112	5.0
0.100	113	113	5.0

TABLE VII

Example of Calculation of Heat Production by Indirect Calorimetry
(Holstein cow 970 lb., 4th day of fast, Jan. 5, 1940)

A. From Carbon and Nitrogen Balance	
Nitrogen lost in urine (measured).........................	53 g.
Protein catabolized (53×6.25)...........................	*331* g.
Carbon in catabolized protein (331×0.52)...............	172 g.
Carbon from protein in urine (calc. as urea) (0.43×53)....	23 g.
Carbon from protein in respiration ($172 - 23$).............	149 g.
Carbon lost as CO_2 in respiration (1275 l. $CO_2 \times 0.532$).....	697 g.
Carbon from catabolized fat ($697-149$)...................	548 g.
Fat catabolized ($548/0.765$)...............................	*716* g.
Heat production from protein (331×4.8).................	1589 kcal.
from fat (716×9.5)......................	6802 kcal.
Heat production per day (from carbon and nitrogen balances).	8391 kcal.

B. From Oxygen Consumption
Oxygen consumed per day: 1730 liters ($0\,°C.$, 760 mm., dry)
Heat production per day (from oxygen consumption) $1730 \times 4.7 = 8132$ kcal.

fat and carbohydrate an animal has catabolized, and in turn what the
caloric equivalent per liter oxygen consumed and carbon dioxide pro-
duced are for various "nitrogen-free R.Q." values. Method and
results of such calculations are given in Table VI.

One may estimate the heat production in kilocalories of an animal
in postabsorptive condition, catabolizing mainly fat, as 4.7 times the

number of liters of oxygen consumed.   This is in line with Thornton's
rule, according to which the heat of combustion of organic com-
pounds, ΔH, is equal to 53 kilocalories *times* the number of gram
atoms of oxygen used in the combustion (*32*).   The calculation on the
basis of the nitrogen and carbon balance is illustrated by the example,
in Table VII, of a fasting Holstein cow, and compared with the ap-
proximate estimate of 4.7 kcal. per liter oxygen consumed.

D. LIMITATIONS AND ERRORS

1. Instrumental Error

A critical discussion of the instrumental errors in calorimetry is
found in White's book (*33*).   White writes: "It is more important
to be assured that large errors are absent than to have very small
values for the small ones."   This is common sense but seems to be
forgotten rather often.   To take great pains in reducing errors from
heat leakage by conduction, convection, and radiation to less than
0.01% of the result, when evaporation may cause errors of 1%, would
not be reasonable.

In applying a generally good technique, relatively large errors
may occur by "blunders or unnoted accidents" (*33*).   Such errors
are detected, as a rule, by repetition of the experiment.   If two re-
sults of the supposedly same process are relatively far apart, one
makes a third measurement; if the result then is close to one of the
previous measurements, one regards the mean between the two similar
results as the correct result and discards the one that differs as the
outcome of a mistake.   This procedure is open to criticism in cases
when the accident cannot be verified.   In that case more measure-
ments are desirable to make sure that one is justified in discarding
the result with the greater deviation.

Also in the absence of accidents or blunders, repetition leads to
averages that have a smaller error than a single measurement, but
this decrease in the error of the mean is proportional to the square
root of the number of measurements.   To reduce the error of a mean
to one-tenth, one has to make 100 times as many measurements.
Repetition therefore, is not a very effective method for reducing er-
rors.

Increase in accuracy is more effectively accomplished by analyz-
ing the errors, finding the causes of as much of the error as possible,

and thus reducing the random error, which is that part of the error for which the causes are unknown. If, for example, the mean error of a heat determination amounts to 0.3% of the result, one may reduce the error of a mean to 0.1% by running nine measurements. If, instead, one finds that $\frac{2}{3}$ of the error originates from uncertainties of water evaporation and discovers a method of avoiding this evaporation error, a single measurement is as reliable—or has as small a standard error—as the mean of nine measurements before the particular source of error was discovered and corrected.

2. Sampling Error

Two determinations of the heat of combustion of a sample of coal or food or similar mixed substances should agree within about 0.5% (Swiss Institute of Fuel Investigations, personal communication). The major part of the error of this determination is usually caused by the lack in uniformity of the samples and only a small part by the inaccuracy of the calorimeter measurement *per se*.

3. Biological Variation

White wrote in 1928 as follows: "In most cases today the desirable calorimetric precision is less than might be readily attained but is all that is worthwhile in view of errors elsewhere in the experiment."

Microcalorimeters can now be made with no more than 0.2 to 0.5% error (*17*). In combustion calorimetry even a crude apparatus consisting of an open can with water for the bomb, with a hand stirrer, and thermometer will often give a precision of 1% (*33*). The Russel Sage respiration calorimeter for man has an error for oxygen consumption of 1.6% and for heat flow of 0.9% (*34*). In alcohol tests the respiration chamber for indirect calorimetry with cows at California showed a standard error of 1.2%. The biological variation involved in the application of these instruments is considerably larger, as a rule, than the instrumental error.

Following White (*33*), one may consider an instrumental error negligible when it does not exceed one-tenth of the biological variation. A coefficient of variation of 4% was found by us for the four-hour measurement of the rate of oxygen consumption of normal rats. For half hour measurements this coefficient amounted to 9%. This is the variability that results when the inherent variance between rats

is subtracted from the total variance. To make instrumental errors negligible in this case they should not exceed 0.4 and 0.9%, respectively.

For human basal metabolism Berkson and Boothby (35) noted a mean coefficient of variation of 4%. With the standardization possible in measurements with human beings, one is thus justified in decreasing the instrumental error of a respiration apparatus to 0.4%.

If the result is the difference or sum of various figures, the contribution of the errors of the measurement to the error of the result is given by the following formula:

$$\epsilon_R^2 = \Sigma \epsilon^2 \tag{19}$$

where ϵ_R is the error of result attributable to summated error of measurement, and ϵ the error of each measurement.

Even reasonably small relative errors may of course become very serious when the result is a small difference of large items. If, for example, the heat of combustion of food has an error of 2% and if the gain in body substance amounts to only 10% of the energy in the food, the error of the food analysis alone causes an error of 20% of the result.

In dealing with the accuracy of the result, one should keep in mind the degree of accuracy warranted by the units in which the result is expressed. If, for example, a 20% error is inherent in the definition of an animal's surface area, not much is gained by improving the accuracy of the metabolic rate 1%, and then expressing the result per unit of the uncertain surface area.

4. Errors in Interpretation

In many cases, as in the one mentioned above, the errors of measurement are far outweighed by mistakes of interpretation. Investigators in the field of animal bioenergetics rather frequently overlook multiple correlation in their systems. One deduced, for example, from his measurements, a relation between body size and metabolic rate of rats, forgetting that the heavier animals were also the older ones and that what he described as an effect of body size may have been an effect of age.

5. Limitations

Berthelot's hope of using heat of reaction as a direct index of chemical affinity did not materialize. The heat of reaction does not permit

one to predict the direction a chemical process is likely to take. The "driving force" of a chemical reaction (11) is measured by the decrease in free energy. A given increase in free energy indicates for a given process how much work, or electrical or radiant energy, is *at least* necessary to make the process take place. The knowledge of free energy, however, does not permit the prediction of the work actually required to produce a given reaction, or the work that can actually be gained from a reaction. The free energy also does not predict the rate at which a given reaction will proceed. The fact that glucose has a heat of combustion of 673 kcal. per mole tells us only that so much heat is developed *when* 1 mole of glucose combines with oxygen to form carbon dioxide and water. It does not tell us *whether* in a mixture of glucose, carbon dioxide, and water, more glucose and oxygen will be found after a given time with heat absorption, or more carbon dioxide and water with heat production.

That the combustion of glucose is accompanied by a decrease of free energy of 688 kcal. indicates a "driving force" in the direction of oxidation rather than synthesis of glucose. This knowledge, however, does not enable us to predict whether or not, in a given mixture of glucose, oxygen, carbon dioxide, and water, combustion at a measureable rate will take place at all. That question involves the kinetics of catalytic processes, especially enzyme reactions, and is beyond the limits of calorimetry and even thermodynamics.

E. ACCOMPLISHMENT OF CALORIMETRY AND OUTLOOK

1. Practical Importance of Heat Transfer

Heat transfer is so universal that hardly any branch of modern industry could function without the great amount of knowledge on thermal processes that has accumulated during the last two centuries. The importance of calorimetry goes even beyond factories and machines. The planning of houses and clothing, the methods of building dams and bridges are among the activities that depend on calorimetric data, many of them available in handbooks. The use of such data would be simplified if the engineers among the English-speaking peoples were to follow the example of the engineers of the rest of the world in regarding engineering as applied physics. This attitude would lead to a unified terminology in all modern calorimetric and

thermodynamic tables, and the terms "grains of moisture," "foot pounds," and "B.T.U." could then be stored in historical museums.

2. Calorimetry as Part of Bioenergetics

Lavoisier demonstrated that animal heat is produced by the combustion of organic compounds in the animal body. This was the first major triumph of biocalorimetry—life may be regarded and studied as a combustion process. The establishment of such a unifying principle in our knowledge of nature is at least as important as the discovery of new phenomena.

Half a century after Lavoisier's achievement, calorimetry led to an even greater generalization. Robert Mayer noted that work could produce heat or that heat could produce work, and he conceived the idea that heat and work were different manifestations of a common agent that could neither be produced nor destroyed—energy. From the difference in the heat capacity of gases at constant pressure and at constant volume Mayer calculated the caloric equivalent of work.

Simultaneous measurements of a dog's heat loss, urinary nitrogen excretion, and carbon dioxide production, carried out by Rubner (1894), confirmed Lavoisier's theory concerning the origin of animal heat, and proved that the fundamental law of thermochemistry is applicable to animal metabolism.

That human work is performed in accordance with the law of conservation of energy was demonstrated by Atwater and Benedict (1899–1903) with the Atwater-Rosa respiration calorimeter. Among recent accomplishments in biocalorimetry the work of Hill and coworkers on the heat production in muscle and nerve fibers already mentioned (27) is particularly famous. The investigation of Borsook and Winegarden (36) also comes to mind. These biochemists applied the second law of thermodynamics to calculate the minimum osmotic work involved in the excretion of urine. They noted that the kidney has a great capacity for work but a low energetic efficiency of 1–2%.

3. Outlook

Energy transformation is one of the most general and fundamental characteristics of life. Biocalorimetry will therefore remain an essential part of physiology. Nutrition at present is dominated by the tremendous successes of vitamin chemistry. In some courses and text-

books on nutrition, therefore, energy relationships are dealt with as a subject of mere historic interest. In these extreme cases, as DuBois puts it, nutrition becomes synonymous with dietetics. When the period of predominant expansion is superseded by a period of intensification of knowledge in nutrition, interest in energy relations will probably be renewed. Such a tendency is already noticeable since vitamins and their functions are linked to enzymes, and the importance of a considerable part of these enzymes is best understood as that of catalysts required in the metabolic processes that provide the energy necessary in all manifestations of life.

References

1. Mach, E., *Prinzipien der Wärmelehre*. Barth, Leipzig, 1919.
2. Maxwell, J. C., *Theory of Heat*. Longmans, Green, London, 1921.
3. Armsby, H. P., and C. R. Moulton, *The Animal as a Converter of Matter and Energy*. Chemical Catalog Co., New York, 1925.
4. Magie, W. F., *Principles of Physics*. Century Co., New York, 1911.
5. *International Critical Tables*, Vol. V, McGraw-Hill, New York, 1929.
6. *Smithsonian Physical Tables*. Smithsonian Inst., Washington, 1933.
7. *Handbook of Physics and Chemistry*. Chemical Rubber Pub. Co., Cleveland, 1946.
8. Sackuhr, O., *A Textbook of Thermochemistry and Thermodynamics*. Macmillan, London, 1917.
9. Rosenthal, I., *Arch. Anat. Physiol., Physiol. Abt.*, **13**, 1 (1889).
10. Burton, A. C., *J. Nutrition*, **9**, 261 (1935).
11. Lewis, G. N., and M. Randall, *Thermodynamics*. McGraw-Hill, New York, 1923.
12. Forsythe, W. E., *Measurement of Radiant Energy*. McGraw-Hill, New York, 1937.
13. Hardy, J. D., *J. Clin. Invest.*, **13**, 593 (1934).
14. Spoehr, H. A., *Photosynthesis*. Chemical Catalog Co., New York, 1926.
15. Saha, M. N., and B. N. Srivastava, *A Treatise on Heat*. Indian Press, Allahabad, 1935.
16. Sturtevant, T. M., "Calorimetry," in *Physical Methods of Organic Chemistry*, 2nd ed., A. Weissberger, ed. Interscience, New York, 1949, Chaps. I and II.
17. Swietoslawski, W., *Microcalorimetry*. Reinhold, New York, 1946.
18. Richet, C., *La chaleur animale*. Alcan, Paris, 1889.
19. Rubner, M., "Die Luftcalorimetrie" in E. Abderhalden, *Handbuch der biologischen Arbeitsmethoden*, Abt. IV, Teil 10. Urban & Schwarzenberg, Berlin, 1926, pp. 819–844.
20. Noyons, A. K., *The Differential Calorimeter*. Fonteyn, Louvain, 1927.

21. Hári, P., "Elektrische Kompensationscalorimetrie," in E. Abderhalden, *Handbuch der biologischen Arbeitsmethoden*, Abt. IV, Teil 10. Urban & Schwarzenberg, Berlin, 1926, pp. 711–754.

22. Capstick, J. W., "Ein Calorimeter für das Arbeiten mit grossen Tieren," in E. Abderhalden, *Handbuch der biologischen Arbeitsmethoden*, Abt. IV, Teil 10. Urban & Schwarzenberg, Berlin, 1926, pp. 793–818.

23. Atwater, W. O., and E. B. Rosa, *U. S. Dept. Agr., Office Expt. Sta. Bull.*, No. 63, 1–94 (1889).

24. Auguet, A. and J. Lefèvre, *Compt. rend. soc. biol.*, **100**, 251 (1929).

25. Murlin, J. R., and A. C. Burton, *J. Nutrition*, **9**, 233 (1935).

26. Whipp, B., *Phil. Mag.*, Series, 7, **18**, 745 (1934).

27. Hill, A. V., *Proc. Roy. Soc. London*, **B111**, 106 (1932).

28. Tian, A., *J. Chem. Phys.*, **20**, 132 (1932).

29. Meyerhof, O., "Mikrocalorimetrie," in E. Abderhalden, *Handbuch der biologischen Arbeitsmethoden*, Abt. IV, Teil 10. Urban & Schwarzenberg, Berlin, 1926, pp. 755–792.

30. J. Pierce Laboratory, *Collected Papers*, Vol. I, 1934–1939.

31. Winslow, C. E. A., and L. Greenburg, *J. Am. Soc. Heating Ventilating Engrs.*, **41**, 149 (1935).

32. Eastman, E. D., and G. K. Rollefson, *Physical Chemistry*. McGraw-Hill, New York, 1947.

33. White, W. P., *The Modern Calorimeter*. Chemical Catalog Co., New York, 1928.

34. Riche, T. A., and G. F. Soderstrom, *Arch. Internal Med.*, **15**, 805 (1915).

35. Berkson, J., and W. M. Boothby, *Am. J. Physiol.*, **121**, 669 (1938).

36. Borsook, H., and H. M. Winegarden, *Proc. Natl. Acad. Sci. U. S.*, **17**, 3 (1931).

QUICK-FREEZING AND THE FREEZING-DRYING PROCESS

Earl W. Flosdorf, *F. J. Stokes Machine Co.*

A. PRINCIPLE OF QUICK-FREEZING

By freezing living organisms and tissues quickly, it is often possible to prevent alteration in their chemical and biological characteristics. Eutectic separation with resulting concentration of components is avoided. This is in contrast with slow freezing, in which two or more phases are obtained, the phase to solidify first consisting of solvent

less rich in solute than the remaining liquid phase. In the case of cellular materials, disruption of cell walls and the like is usually avoided. Protein molecules, which frequently are large and relatively labile, are locked in place as a result of quick-freezing and denaturation does not occur. Viability of microorganisms is maintained. Enzymic changes are retarded if the product is frozen quickly. In a word, quick-freezing is applied for the basic purpose of avoiding change.

The speed of freezing of biological materials depends upon several factors in addition to the temperature of the refrigerant. These additional factors control the speed of freezing by influencing the rate of transfer of the latent heat of fusion from the material being frozen to the refrigerant. Of obvious importance is the extent of the surface exposed to the source of low temperature. The existence of any insulating barrier is a deterrent to rapid freezing, although many substances must be protected by a satisfactory type of container of metal, glass, or other material. Ice itself forms an insulating layer that retards further freezing. Heat transfer at a liquid–liquid interface is more rapid than at a liquid–gas or solid–gas boundary. The effect of the latter factor on quick-freezing has been discussed by Goetz (1), who pointed out that heat transfer from a warm object to liquid air can be slow in spite of the large temperature gradient, owing to the formation of a gaseous layer between the liquid air and the material being frozen. With liquids, agitation or stirring is beneficial.

B. PRINCIPLE OF DEHYDRATING FROZEN MATERIALS

The objectives and principles involved in the dehydration of frozen biological materials have been discussed in detail in a recent book by the author (2). These may be summarized as follows.

First, the temperature is below that at which many labile substances undergo chemical change. This applies to labile components in blood, to viruses and most forms of microorganisms, and to other biologicals and pharmaceuticals.

Second, because of the low temperature, the loss of volatile constituents is minimal. This is particularly important in application to many foods like orange juice and pineapple juice.

Third, since the product is frozen, there is no bubbling or foaming. Thus, changes due to surface action such as the surface denaturation

of proteins, which occurs in drying their solutions even at low liquid temperatures under vacuum, may often be avoided.

Fourth, in most cases, the solute remains evenly dispersed and distributed without undergoing concentration as the frozen solvent sublimes. The remaining dry residue emerges as a highly porous solid framework, which occupies essentially the same total space as the original solution. Hence the final residue is not the fine powder with which the chemist is mostly familiar but consists of a friable interlocking, and spongelike structure. As a result, solubility is extremely rapid and complete. For example, gelatin dried in this manner, from a solution that had to be prepared in the first place by boiling, becomes instantly soluble in cold water.

Fifth, since the molecules of solute are virtually "locked" in position in this way, the tendency for coagulation of even lyophobic sols is minimal. Even though the lipoid constituents of dry blood plasma do not reconstitute perfectly after drying and do produce a slight degree of turbidity, there is far from complete coalescence. The particles are small enough to be safe for intravenous injection and do not cause capillary embolism.

Sixth, during drying the surface of the evaporating frozen ice layer gradually recedes to leave more and more of the highly porous residue of solute exposed. Consequently, "case-hardening" never occurs. A far lower content of moisture may be obtained in the final product without using an excessively high final temperature. Owing to this lower moisture content, a greater degree of stability results than is the case after any other method of drying

Seventh, bacteriological growth and enzymic changes cannot take place under the conditions of freeze-drying. This is important for foods as well as medical products used in parenteral injection. The final fully dried product likewise resists bacterial growth and enzymic action.

Eighth, because of the high vacuum used, in contrast with the degree of vacuum used in ordinary low temperature liquid evaporation, the amount of oxygen present is extremely small, so that even the most readily oxidizable constituents are protected.

A decade ago this dehydration process was largely a laboratory curiosity and most frozen products were stored in a refrigerator to prevent thawing. As a result of recent advances (2) it has been recognized that by establishing proper vacuum conditions for removal of

water vapor and by rapid application of heat to the frozen product, rather than keeping it in an icebox, the process could be made more practical, with resulting improvement in products. Even while being heated, products may be maintained at any desired range of temperatures such as $-10°$ or around $-40°C$. simply by regulating the degree of vacuum and rate of evaporation.

There are two stages in drying by sublimation. In the first, ice is evaporated from a frozen mass. In the second, moisture is removed from the final dry solid to lower the residual content to a minimal level. During the first stage, depending on the particular product, some 98 to 99% of all water is removed. In the second, the residual moisture content is reduced to 0.5% of the final product or less; this represents final removal of 99.95% of the original content of water (assuming 10% solids originally). In the first stage, temperatures are well below $0°C$. Actual temperature varies with the product, as will be discussed later. Upon passing from the first stage to the second the temperature gradually rises and finally reaches that of the room or higher, depending upon whatever final ambient temperature is used.

Basically, during the first stage of drying a maximal rate of evaporation for a given temperature of the frozen product is desirable. To achieve this under some circumstances it may be necessary to heat the frozen product but without causing it to soften or melt. At the same time a maximal rate of flow away from the evaporating surface must be established. To accomplish this rapid flow adequate passageways must be provided for vapor, and this must then be condensed or evacuated efficiently.

C. POTENTIAL APPLICATIONS

1. Quick-Freezing Materials for Storage

Frequently in carrying out an investigation with an unknown biological system, it has not been established whether there are labile components present or not. Manipulations may extend over a matter of days or even months. By utilizing quick-freezing and storage at low temperature at all times when not actually working with preparations, loss of possible labile components is avoided. This type of investigation was encountered in studies with the antigenic structure of *Hemophilus pertussis* (3). Eventually it was found that

agglutinogen is heat stable and freezing overnight or at other times is unnecessary.   Of the two toxins present, one was found to be heat stable and the other extremely labile.   By application of freezing during storage, it was reasonably certain that no possible labile components were being missed as a result of deterioration during the rather extended time required to conduct the investigation.   Another advantage in freezing is the prevention of bacterial and other contamination without the use of preservatives, which are frequently undesirable because of deleterious action.

2. Low Temperature Chemical Manipulations

Cohn (4,5) and co-workers have separated various fractions from human blood plasma under low temperature conditions.   These fractions have been obtained in a high degree of purity and as a result have been used for many specialized medical purposes.   During World War II, fractions were obtained from plasma collected by the American Red Cross; for example, serum albumin in high concentration was used in place of whole blood for osmotic effects, etc.   In this work, Cohn avoided salting out, which is the classical procedure for separation of proteins, and thus avoided dialysis.   Ethanol is used in various concentrations, and, depending upon the relative solubility of various blood derivatives under conditions of varying salt content, pH, and temperature, and with the use of relatively low temperature (0 to $-10°$C.) as well as other means, the proteins are precipitated. By varying these conditions in proper sequence, separation into different fractions is accomplished (5).   The Cohn system is, among other things, an application of low temperature in systems that do not actually freeze in order to carry out precipitations and other manipulations without chemical alteration, such as denaturation of proteins, which proceed at room temperature.

3. Preservation of Laboratory Cultures

Drying from the frozen state has become an extensively used laboratory process during the past ten or fifteen years (2).   Preservation of microorganisms such as bacteria and viruses is widely employed to eliminate continued subculture with possible variation and degrading.   *Neisseria meningitidis* (meningococcus), *N. gonorrheae* (gonococcus), *Hemophilus influenzae, H. pertussis, Eberthella typhosa,* and other similar organisms have been kept viable for a period of

many years; upon reconstitution growth may be started again to produce subcultures of unaltered characteristics.

Influenza virus may be successfully preserved (7) and freeze-drying is used widely for carrying various strains of the virus in many research laboratories. Hoffstadt and Tripi (13) report successful three year preservation of Levaditi and Cutter strains of vaccinia, herpes simplex, laryngotracheitis of fowls, and Rous sarcoma viruses. On the other hand, they found inconstant maintenance of viability of the virus of infectious myxomatosis of rabbits over the three year periods; whether this was because of nonuniform residual content of moisture or other cause was not determined. Their culture of OA strain of Shope's fibroma did not survive the period. Other viruses that have been successfully kept are those causing hog cholera (14), rinderpest, ovine ecthyma (sheep scabs), yellow fever, and various fowl diseases such as laryngotracheitis and fowl pox prepared from chick embryos.

Libby (15) has used freeze-drying in immunochemical studies with tagged antigens involving radiotracers, also with tobacco mosaic virus tagged with radioactive phosphorus (P^{32}). Mice were injected with the tagged virus, 24 hours later the mice were killed in a Dry Ice–acetone bath and 3 to 4 mm. sagittal sections were prepared with a band saw. These sections after freeze-drying were impregnated with paraffin and 1 mm. sections were prepared with a microtome and placed in intimate contact with X-ray film for exposure for varying periods to obtain radioautographs (also see chapter on radioactive isotopes). In this way it is possible to locate regions and organs of greatest concentrations of radioactivity.

Bacteriophage from many organisms has been dried, including that of dysentery (19). Harris and Shackell used drying in a glass desiccator for preservation of rabic brains (9). Rivers and Ward have relied widely on freeze-drying in their work with an intradermal vaccine for Jennerian prophylaxis (10). Siedentopf and Green (11) have reported great success in the preservation of modified canine distemper virus (ferret passage).

4. Preservation of Body Fluids and Miscellaneous Products

As a research tool in the laboratory, freeze-drying offers many possibilities. The porosity and friability of the dehydrated material makes it particularly adaptable to rapid and sterile extraction with

ether, petroleum ether, alcohol, and similar liquid solvents. De-
naturation of the proteins on such extraction is greatly reduced as
compared to extraction of liquid products. Various proteins may be
freeze-dried simply for preservation during the course of an investiga-
tion. For example, many enzymes have been dried without loss in
activity (2). Crystalline beef liver catalase has been dried; the
properties of the dried product compared to those of undried crystal-
line beef liver catalase. Dounce and Howland (23) report that the
dried material is not crystallizable.

Guinea pig serum may be preserved with no loss in complementary
activity over a period of years, even with storage at room tempera-
ture, in contrast with loss of up to 50% in activity on storage over-
night at 5°C. as a liquid. Diagnostic laboratories and research lab-
oratories may gain the benefit of uniform complement obtained from
a pool of blood from a large number of animals even though only
small quantities are used at a time. By preparing such a large pool
and freeze-drying it in small containers, the product is available as
required.

Many other biological products have been dried successfully in the
frozen state with no change in properties and with the final products
having excellent characteristics of solubility. Teague, Galbraith,
Hummel, Williams, and Macy report that the removal of water from
feces, urine, and milk by dehydration in vacuo from the frozen state
permits indefinite preservation of the dry material in an undenatured
form if stored under proper conditions (20). This method of dehy-
dration has many advantages over oven drying. Oven drying, at
70°C. and under, hydrolyzes the soaps in feces, causing exaggerated
values for the free fatty acids and a reduction in the soaps, although
the total free fatty acid plus soap values are the same for both methods
of drying. The nitrogen contents of the Cryochem samples of feces,
urine, and milk approximate those of fresh specimens. In obtaining
energy data by the bomb calorimeter, the Cryochem-dried material
permits greater accuracy in analyses and economy of time and ma-
terials by eliminating one nitrogen determination and the correction
for nitrogen loss in drying.

Bile has been prepared in dry form, so that a stable product may be
stored ready to dissolve for use in obstructive jaundice. It is believed
that the abnormal bleeding that occurs is due to the failure of the
absorption of vitamin K in the absence of bile in the intestines (21).

Farr and Hiller have reported on the successful drying of hemo-

globin (*22*).   Further details will be found in Section E5 on the influence of oxygen on the life of frozen-dried preparations.

5. Histological and Cytological Preparations

Freeze-drying has been employed extensively for many years by histologists and cytologists.   The technique as practiced is quite different from the method used today for biologicals, pharmaceuticals, microorganisms, and other products discussed.   The procedure is much slower and the products are maintained at lower temperatures without heating.   It is for these reasons, as well as the fact that tissues are generally more difficult to dry than liquids, that the process is more time consuming and requires more elaborate equipment.

It was nearly sixty years ago that Altmann (*17*) stated that biological materials could be dried without shrinkage at a temperature of $-20°C$.   This was published in connection with his cytological work on bioblasts.   Forty years later Gersh (*42*) described equipment for fixation that involved the use of liquid ammonia to provide the low temperature and a diffusion pump and phosphorus pentoxide for removal of the moisture.   With this equipment Gersh was able to confirm the value of the procedure for fixation of some organs and tissues (skin, cartilage, smooth muscle, liver, pancreas).   Gersh applied the technique successfully to the investigation of the excretion of uric acid and ferricyanide, the chemical nature of intracellular constituents, and of the intracellular distribution of glycogen.   Gersh also used the method for the preservation of vitally, or supravitally, stained preparations that fade in fixation or are incapable of satisfactory preservation by other means.   He found that fixation of organs other than those just mentioned was less satisfactory and that central nervous system material is exceedingly poorly fixed.   Results of very interesting investigations on these subjects were then published by Bensley, Gersh, Stieglitz, Hoerr, and others during the next three years (*43*).   Goodspeed, Uber, and Avery reported on interesting applications in plant cytology.   They studied particularly the character of fixation and reported on chromosome structure in *Lilium longiflorum* (*44*).

6. Commercial Processing of Heat-Sensitive Materials

During the past few years, freeze-drying has been applied to the production of many medicinal products on a scale that fifteen years

ago might have seemed fantastic.   There are some commercial plants
capable of removing water from the frozen state in terms of thousands
of liters per day.   During the war, all the plasma prepared from
blood collected by the American Red Cross for our armed forces was
dried from the frozen state for distribution.   Some of the fractions
prepared from plasma by the Cohn procedure are freeze-dried.
Convalescent human serum, antitoxins and other antisera, bacterial
and viral vaccines, penicillin, streptomycin, parenteral vitamin prepa-
rations, and other products are dried by this process and are available
commercially.   The principles involved in large scale freeze-drying
are the same as on a laboratory scale discussed in this chapter but,
of course, the manner of application varies considerably.   This is be-
yond the scope of the present book and fuller discussion may be found
elsewhere (2).

D. EQUIPMENT FOR FREEZING AND DEHYDRATING

1. Sources of Low Temperature

For quick-freezing, Dry Ice (solid carbon dioxide) suspended in a
bath of organic solvent such as ethanol, acetone, or the like is the
simplest and most readily available refrigerant for low temperature.
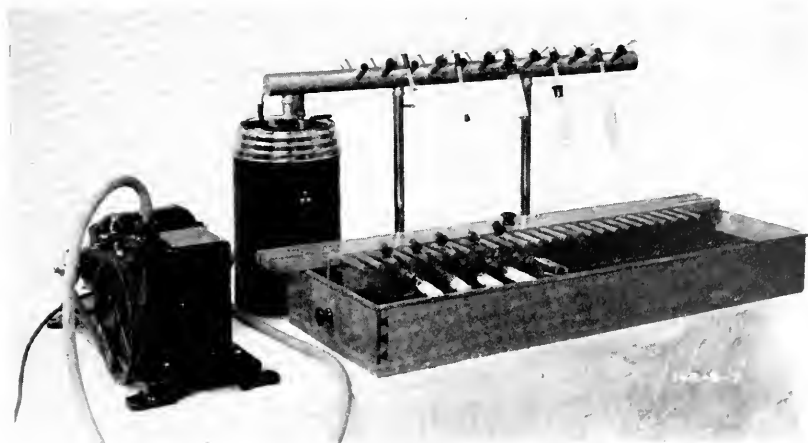
Fig. 1.   Freezing-drying equipment employing solid carbon dioxide (Dry Ice)
for general research.   (Courtesy *F. J. Stokes Machine Co.*)

Liquid air or nitrogen may be used for still lower temperature. Dry Ice in large Thermos bottles is also simple and convenient for storage. The ice cube compartment of household refrigerators is suitable for storage of small quantities of product and even for freezing when the amounts involved are small. For more rapid freezing than otherwise would be obtained by use of the ice cube compartment, a bath of
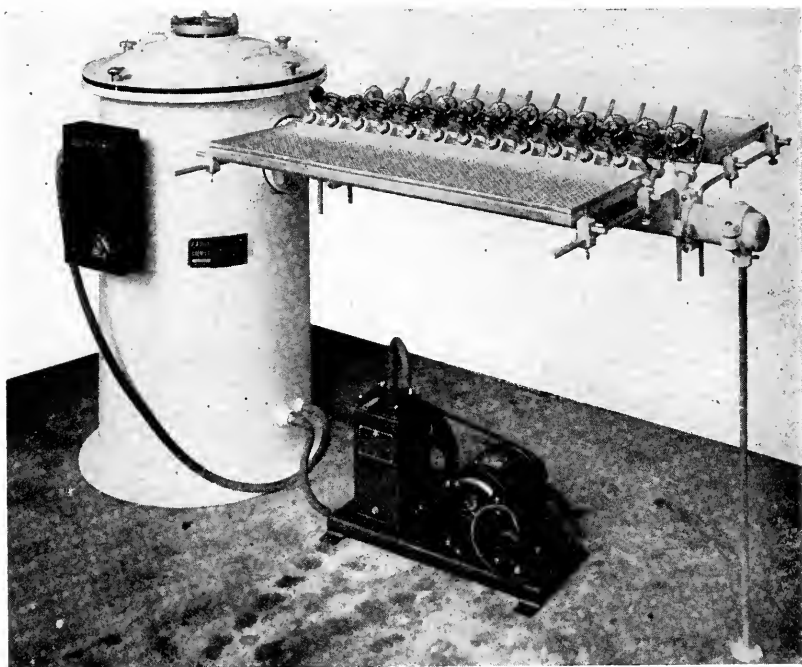
Fig. 2. Cryochem type of equipment for desiccating from the frozen state. Regenerable desiccant comprising mostly calcium sulfate (Drierite) with a small amount of silica gel (for small amount of vapors other than water) contained in baskets within the white tank. Manifold is shown with outlets to which individual containers for drying are attached. At the end of drying, the glass tubes attaching the containers to these outlets are sealed by fusion with a flame. (Courtesy *F. J. Stokes Machine Co.*)

alcohol may be chilled to a low temperature beforehand. Then by immersing a glass container of the product in the alcohol a complete and rapid freezing may be obtained because of the faster transfer of heat to a liquid. When larger amounts of material are involved, some of the currently available home freezers are suitable. The only technical skill required is good judgment in the matter of rotat-

ing the containers and other manipulations to achieve either rapid
or slow transfer of heat depending upon which type of freezing is de-
sired.

2. Apparatus for Dehydrating Frozen Material

Many of these have been described but all operate basically on
similar principles.   Either low temperature condensation or chemical
removal of water vapor under high vacuum is used for laboratory type
equipment.   Typical apparatus is illustrated in Figures 1 and 2.   A
full bibliography of the various published types of apparatus will be
found elsewhere (2).   Packer and Scott (46) have reported on a sim-
plified cryostat for the dehydration of frozen tissues.   This was an al-
most entirely glass apparatus permitting a dehydration temperature
for the product as low as −70°C.   They claim low cost of operation;
Dry Ice was used for the refrigerant.   Suitable lyophile equipment
(24) utilizing Dry Ice condensation and freezing, and Cryochem type
of equipment using a regenerable chemical desiccant (8) have been de-
scribed.   Detailed consideration of various principles employed in
applying these methods to medical products was published in 1945
(25).   Laboratory apparatus is available commercially in a wide
variety of sizes and types (48).

E. TECHNICAL OPERATIONS AND LIMITATIONS

1. Criteria for Quality of Preservation

As far as is known, many frozen-dried preparations such as blood
plasma, penicillin, antitoxins, convalescent serum, some species of
microorganisms, and other products have virtually an indefinite life.
Naturally, time must elapse before the full extent of preservation can
be established for any given product.   Accelerated testing may be
carried out by use of higher temperatures such as the rates of deterio-
ration at 20, 50, and 100°C.   Higher temperature, however, is not
always an accurate index of what may be expected at ordinary tem-
peratures.   There may be a critical temperature level above which
a given product will be almost instantly destroyed.   For this reason,
every preparation must be considered individually.

The original properties of some materials may be destroyed in the
process.   For example, multicellular organisms are killed.   Spiro-
chetes and similar organisms do not survive freeze-drying.   In pre-

paring fibrinogen and prothrombin, there may be inactivation of the
product as a result of loss of carbon dioxide, thus producing alkalinity.
This inactivation may be prevented, however, by buffering the orig-
inal solution quite strongly or by acidulating just before drying.
Lyophilic colloids are well preserved, but lyophobic colloids lose their
original properties more or less in proportion to the extent of their
lyophobic characteristics.

The complementary activity of guinea pig serum is one of the bet-
ter examples of a product with a limited life even though its duration
is extended manyfold as a result of freeze-drying.   Refrigeration tests
carried out as long as five years from the date of preparation show
little or no loss, but at room temperature the potency begins to fall
off before this time.   A similar situation exists with viruses.   The
final residual content of moisture in the product is of utmost impor-
tance in many cases in determining the life of the preparation.   In the
case of products in the intermediate class, which have a limited degree
of preservation, the residual moisture is usually more critical than
otherwise.   In all cases it is well to try to obtain the final content of
moisture in the range of 0.5 to 1% for maximal life.

Concentrated globulin preparations dissolve very slowly in re-
constitution.   Instead of being a matter of minutes, the time may be
extended to one-half hour or more.   In this case, however, there is no
loss in biological activity after the product has become dissolved com-
pletely.

With regard to the percentage of viable cells remaining in cul-
tures of bacteria after desiccation there is wide variation, depending
upon the species and also the particular strain;   this frequently may
be as little as 5%, but, since the surviving cells keep well, there is
adequate survival for the purpose of maintenance of stock cultures.

With viruses, there is usually less activity after processing, but,
as with bacteria, the final desiccated product is well stabilized and no
further loss occurs during storage (7).   Wooley (16) has reported
that lymphocytic choriomeningitis and St. Louis encephalitis after
freeze-drying had been preserved for 378 and 833 days, respectively,
tests for longer periods not having been made.

2. Optimum Dehydrating Conditions

Cultures and Miscellaneous Materials.   The actual degree of
vacuum required is a function of the temperature of the material

undergoing desiccation.   It is pressure of water that is important and this is a question of the vapor pressure of the ice at the temperature of the product being dried.   Air must be removed to a pressure of 100 to 200 μ of mercury and in efficient commercial installations to 50 μ.   Penicillin is usually dried at about $-30°$C., where the pressure is 300 μ.   For blood plasma, $-25°$ or 0.5 mm. of mercury pressure is suitable and serum can be dried at $-40°$ or 2 mm.   These conditions must be determined experimentally for each material.   A McLeod gage can be used for measuring total pressure including water vapor, but it must be suitably trapped (2).   Time required for desiccation depends upon the quantity in the container and depth of the frozen layer.   Plasma in 300 ml. quantities requires twenty to forty hours. Cultures in the small quantities usually dried (0.1 to 0.5 ml.) require only five or six hours, but usual practice is to allow overnight drying for convenience.

Viruses are in general more difficult to dry than other substances and a lower temperature must be maintained during dehydration than is frequently necessary with many other products, preferably well below $-20°$C. (8).   Some bacterial and virus cultures undergo as much as 95% loss in percentage viability, whereas in certain other cultures as much as 90% survival is obtainable.   The menstruum is important as well as the rate of drying in obtaining maximum survival.   Delimiting of optimal conditions for the many organisms to which freeze-drying is applied has been carried out for only a relatively few.   In many cases, as for the maintenance of a library of stock cultures without the necessity for continual subculturing or animal passage, extended partial loss of viability is not of serious consequence.   Particularly in the case of live vaccines the percentage of survival is very important and it is here that most work has been centered (2).

Histological and Cytological Preparations.   In the application of the freezing-drying method to animal and plant cytology (18,45), tissue is frozen at a very low temperature, such as in liquid air, which results in killing the tissue.   Then it is dried at a temperature sufficiently low to avoid appreciable diffusion or displacement of cell constituents.

Hoerr (45) and other workers (1) have attempted to improve the quality of tissue fixation by resorting to more rapid freezing of the tissue by means of pentane chilled to $-131°$C., or isopentane chilled to $-195°$.   Dehydration was carried out quite slowly with the prod-

uct at temperatures below $-30°$, permitting the tissue to warm very slowly from the freezing temperature to the dehydration temperature. Paraffin was not used for embedding in microchemical studies. The lower temperature by-passed the greatest objection to the technique thus far for cytological studies by making the size of the ice crystals smaller with rapid freezing so as to reduce the artifacts to below the resolving power of the microscope. Thus we see that the entire trend is one of lower and lower temperature for the product.

In Scott's opinion (47) a dehydrating temperature of $-20°$ does not give particularly good results when one is interested in the electrolytes found in tissues. The reason is, of course, that the eutectic point of many of the salts normally found in tissue is considerably lower than $-20°C$. Consequently, it is impossible to keep them in a salt–ice equilibrium at $-20°$.

The theoretical temperatures of dehydration run as low as $-54.9°$. Even lower temperatures would not be out of place. But for practical reasons it is very difficult to dehydrate tissues in these extremely low temperature ranges. The vapor pressure of ice becomes very low and an adequate pressure differential is difficult to obtain. The dehydration is consequently slowed down tremendously. It would take about six weeks or so to dehydrate 20 g. of ordinary tissue at $-65°C$. Consequently, Scott (47) set a drying temperature of $-32.5°$ and states that this is not theoretically correct but that it is practical and feasible. In general, Scott found that it gave excellent results.

3. Production Scale Processing

Because of the larger quantities that must be handled on a production scale basis, and in order to carry out freeze-drying at a commercially economical cost, the equipment often is considerably complex. Mechanical refrigeration for production of low temperatures is employed. The services of skilled mechanics as well as production laboratory personnel are required. Many of the products are for parenteral use and asepsis must be maintained, so that much attention must be given to the handling of such products in a sterile fashion. It is not always possible to carry on manipulations in a closed system, therefore the workers must use surgical technique with face masks and the other usual precautions. Filtered, dehumidified air should be circulated and sterile lamps widely distributed around the working areas.

4. Role of Protective Proteins

It has been found that a protective protein greatly prolongs the keeping qualities of the organisms, as well as greatly increasing the percentage viability. Serum or plasma may be used but first must be inactivated to remove bactericidal activity. Sterile skim milk (autoclaved) provides a most satisfactory medium (6). Cultures grown in liquid medium may be concentrated by centrifugation and then resuspended in milk. Organisms grown on a solid medium may be harvested by scraping into saline or milk. If saline is used the suspension is then added to an equal volume of sterile skim milk. This mixture is then distributed into small containers. Usually large numbers of containers of culture are dried at a single time. Small quantities per container are ordinarily adequate, as little as 0.05 ml. being satisfactory. It is highly important, however, to use all-glass containers rather than containers that carry a rubber stopper exposed to the atmosphere. Otherwise, diffusion of water vapor through the stopper will quickly raise the moisture content of the culture to a point where viability of many species is not maintained. Microorganisms preserved in such small amounts are especially sensitive to this condition, hence the necessity of using containers that may be sealed by fusion of glass.

5. Oxidation as a Factor in Preservation

It is known that some substances after drying from the frozen state may not retain activity in the presence of oxygen of the air, even though completely dry and hermetically sealed. Typical of such products are those of high lipide content. In such a case either sealing under original vacuum or under an inert gas such as nitrogen or argon is necessary. Siedentopf and Green (11) report that this is the case with their distemper virus. It should be pointed out that release of original vacuum with dry air and subsequent evacuation or replacement of air with inert gas does not operate satisfactorily in all cases, since oxygen may be absorbed by the highly porous solid matter and cannot be readily swept out for removal. Also, for distemper virus the nitrogen must be purified by passage over hot copper (12).

Farr and Hiller (22) found that application of the method of freeze-drying to oxygenated hemoglobin solutions resulted in preparations in which the hemoglobin had lost 25 to 30% of its oxygen-binding capacity, by change to methemoglobin. However, when hemo-

globin solutions were first deoxygenated by repeated evacuation of all gases, so that over 99% of the oxyhemoglobin was changed to reduced hemoglobin, the reduced solutions could then be frozen and dried in ampules and the dried hemoglobin kept *in vacuo* for months without methemoglobin formation. In redissolving the reduced hemoglobin it was necessary to prevent even momentary access of atmospheric oxygen to the dried material before it was dissolved or methemoglobin was formed. After the reduced hemoglobin was in solution oxygenation did not inactivate it and the solution was stable in air. At 4°C. the solution could be kept several weeks without significant change.

6. Storage and Use of Products

During storage of dried cultures ordinary refrigeration at about 5 to 8°C. is used. This is not essential in all cases, but carrying out systematic studies of the effect of temperature on storage of the many hundreds of strains and different species of organisms of interest is almost prohibitive. Accordingly, it is safest to rely upon refrigeration, particularly since the volume of such samples is so small.

For removing the dry organisms, the stem of the all-glass container may be scratched with a file in the usual fashion. Before breaking the glass tip, the container is wrapped with a cloth impregnated with antiseptic to prevent the spreading of dry organisms as air rushes in. Excess antiseptic should first be squeezed out of the cloth, however, to avoid sucking the liquid into the container when the vacuum is broken. Sterile water is then added. With a small loop, the proper culture medium, either liquid or solid depending upon circumstances, is inoculated. A somewhat longer incubation period may be required to obtain satisfactory growth of the first culture generation.

F. SOME OUTSTANDING ACCOMPLISHMENTS

1. Virus Research and Distribution of Vaccines

Quick-freezing for storage, and freeze-drying for preservation of viruses in distribution as viral vaccines has been employed with success. Freeze-drying has been applied to vaccinia virus, rabies (*9*), canine distemper (*11*), influenza, herpes simplex, laryngotracheitis of fowls, Rous sarcoma, hog cholera, rinderpest, yellow fever, and other viral vaccines, some of them for practical distribution on a fairly

wide scale.   One of the earliest applications of freeze-drying was in the control of rinderpest (cattle plague) in Africa.   The mortality for this disease is high, sometimes reaching 50 to 75%.   In preparation of the virus the blood of infected animals is laked and centrifuged. This separates the leucocytes with which the live virus is associated. Without drying, the virus remains viable and effective as an immunizing agent for a matter of a few weeks only.   However, it survives freeze-drying well and has its life extended to months and years. Cattle are immunized by the simultaneous inoculation of living virus and immune serum, this being similar to the practice of immunizing swine against hog cholera.

The viral vaccine for prevention of canine distemper has been produced in a small way commercially.   Various fowl viruses are produced by desiccation.   Vaccine for the prevention of yellow fever in man has been successfully produced on a rather large scale.   In experimental work the virus of influenza has been freeze-dried.   Vaccine for intradermal injection for immunizing against smallpox by Jennerian prophylaxis (10) has all been distributed after freeze-drying. There is some question about the efficacy of this method of immunization but this has no bearing on freeze-drying *per se* (being related to the nature of the material itself and the method of carrying out immunization).

Munce and Reichel (14) have shown that hog cholera virus of blood origin, when desiccated under high vacuum and stored *in vacuo* in flame-sealed ampules, remained infective after exposure to a temperature of 60°C. for 96 hours.   At 37°, the infectivity was maintained for 328 days.   Phenolized liquid virus from the same mixture was noninfective after exposure to the higher temperature for only five hours.   At 37° the period of infectivity of the dried preparation was approximately 23 times as long as for the phenolized liquid virus. After storage at 20°, the dried virus was still infective in the authors' last test, which was conducted after a storage period of 1125 days— 12 times as long as for the corresponding phenolized liquid virus.   It may be pointed out that hog cholera virus, as commercially available, consists of the phenolized, defibrinated blood of artificially infected swine undergoing an attack of acute hog cholera at the time their virus-laden blood is collected.   This virus preparation is frequently referred to as "simultaneous hog cholera virus" because it is used simultaneously with anti-hog cholera serum in the immunization of swine against hog cholera.   The animals would not survive the injec-

tion of the live virus were they not protected simultaneously with the serum.   When used in conjunction with anti-hog cholera serum, the virus possessing high virulence and the proper antigenic properties stimulates a strong, active immunity of lasting duration.   The product when dispensed as a liquid must be preserved with phenol in the amount of 0.5%.   This agent exerts a virucidal action that gradually reduces the viability until it is no longer infective.   For this reason, the United States Bureau of Animal Industry has assigned an expiration dating of only ninety days for the phenolized liquid from the date of the *production* (not sale) of the product.   Accordingly, the results with the freeze-dried product are of particular importance, not only in extending the actual dating, but also in assuring that the product when used will not have undergone partial deterioration.

Munce and Reichel have also reported that the freeze-dried samples maintained their infectivity for a longer period after freeze-drying when stored in flame-sealed ampules than in rubber-stoppered bottles, in both cases the containers being evacuated.

Although unpreserved liquid virus maintains its infectivity for a longer period than phenolized liquid virus from the same mixture and stored at the same temperature, governmental regulations require the use of a suitable preservative, for obvious reasons.   Accordingly, the results of significance are those comparing the longevity of the freeze-dried product with the phenolized liquid virus.

2. Bacterial Preservation

Freeze-drying of stock cultures to avoid continual laboratory subculture is a routine practice in most of the more important bacteriological laboratories today.   The method not only results in saving of time and labor, but it avoids the variations that occur with so many organisms after continued laboratory subculturing.

Hammer applied desiccation for the preservation of numerous viable bacteria (26).   Rogers used the method successfully for lactic-acid-forming bacilli (27).   Swift has used this method of desiccation extensively in his work with streptococci and pneumococci (28). Siler and associates at the Army Medical School have maintained their cultures of the now well known S-58 virulent *Eberthella typhosa* without dissociation over a period of many years (29).   Preservation of this strain in this fashion made it possible for Siler and his group to embark on a program of research lasting over a period of years.

There would have been little justification for starting without the assurance that at the end they would still be dealing with an organism of unchanged characteristics.   Freeze-drying made this possible. As a result of this work carried on by Siler and associates, a far better vaccine for preventing typhoid fever was produced and was available for use in World War II.   This is another major, although indirect, contribution that freeze-drying made to the great success of American military medicine during the war.   Welch, Borman, and Mickle reported successful application to *Klebsiella pneumoniae* in unaltered form (*30*).

Flosdorf and Kimball reported on extensive use of freeze-drying with *Hemophilus pertussis* (*6*).   In this case, agglutinin absorption was used to demonstrate that no dissociation occurs as a result of drying.   The fundamental importance in antibacterial immunity of the combination of agglutinins with surface antigens of bacteria is well established.   With nonflagellated organisms the use of such surface reactions as agglutination or phagocytosis therefore provides distinguishing methods of assay for effective surface reactants in either serum or antigen.   Complement fixation and precipitin testing with soluble antigens are of diagnostic value but do not distinguish surface antigen-antibody combination from phenomena involving other antigens.

Appleman and Sears (*31*) reported that legume nodule bacteria (*Rhizobium leguminosarum*) retain viability and their ability to nodulate plant hosts and to fix nitrogen after four years of storage without loss.   Bacteria tested were isolated originally from alfalfa, lespedeza, cowpea, pea, soybean, vetch, crown vetch, and clover host plants. The cultures were grown on asparagus mannitol medium and then emulsified in sterile water for drying.

Freeze-dried viable cultures of molds and bacteria are now available commercially in the dairy industry as "starters" for first propagation in production of cottage cheese, buttermilk, butter acidophilus, Bulgarian milk, and yoghurt.

3. Preservation of Miscellaneous Products

As has been discussed earlier in this chapter, blood plasma was preserved on a very large scale during World War II.   It represented the largest application of freeze-drying until that time.   Plants were erected throughout the country for the purpose.   The method was also employed as part of the fractionating procedure in the Cohn

method of separating the various components of blood plasma. Some of the fractions were freeze-dried in final form for actual distribution. Penicillin, streptomycin, and other antibiotics have been freeze-dried for distribution. Freeze-drying has been applied to the distribution of many other products because of lability and because at the same time there is greater ease of control of sterility as well as an increased rate of solubility. This latter consideration is particularly important where an ampule of the product must be opened by a physician and dissolved at a temperature not lower than that of the room and not above that of the body and made ready for immediate injection. Rapid solubility frequently determines the suitability of a product for clinical use. Freeze-dried preparations of hormones, parenteral vitamin B preparations, enzymes, and other products are finding continually wide: distribution.

Highly purified tuberculin protein satisfactory for skin testing in control of tuberculosis has been dried successfully by lyophilization. Seibert, who has contributed more than anyone else to the chemistry of tuberculin, has studied the effects of freeze-drying and various procedures of applying it. She has found no change in this protein material detectable by some of the most sensitive means available today (32). Langner and associates have applied freeze-drying to organisms for extraction and to extracts obtained from them in connection with studies in allergy (33–35). Casals has applied freezing and drying to various antigens of a nonvirulent nature to be used in complement fixation tests with central nervous system virus infections. These have been frozen and dried and distributed to various laboratories for use in hospital diagnostic work (36).

Smallpox and BCG vaccines are dried in the U. S. S. R. with excellent results for shipment to remote localities (37). Hetherington has used lyophilized serum, embryo juice, and plasma media for tissue culture (38). An adequate supply of material can be kept on hand for continued use to avoid considerable routine labor and at the same time provide a constant medium for use in a long series of comparative experiments. The product has been found excellent in connection with the growth of cardiac explants for white mice. Souter and Kark have produced a stable thromboplastin suitable for use in the Quick prothrombin test (39). Material ready for immediate use with the addition of distilled water permits the performance of the test by the physician in his own office and makes it possible to carry out the test with greater ease anywhere.

Considerable exhaustive investigation is being carried on in the field of food products, such as orange juice, raw ground beef, soluble coffee, and other products, with the objective of distribution without refrigeration and the development of reconstituted products having nutritional characteristics as well as palatability equivalent to the fresh product (2,40,41).

References

1. Goetz, A., and S. S. Goetz, *Proc. Am. Phil. Soc.*, **79**, 361 (1938); *J. Applied Phys.*, **9**, 718 (1938).
2. Flosdorf, E. W., *Freeze-Drying*.   Reinhold, New York, 1949.
3. Flosdorf, E. W., and A. C. Kimball, *J. Immunol.*, **39**, 475 (1940).
4. Cohn, E. J., J. L. Oncley, L. E. Strong, W. L. Hughes, Jr., and S. H. Armstrong, Jr., *J. Clin. Investigations*, **23**, 417 (1944).
5. Cohn, E. J., L. E. Strong, W. L. Hughes, Jr., D. J. Mulford, J. N. Ashworth, M. Melin, and H. L. Taylor, *J. Am. Chem. Soc.*, **68,** 459 (1946).
6. Flosdorf, E. W., and A. C. Kimball, *J. Bact.*, **39**, 255 (1940).
7. Scherp, H. W., E. W. Flosdorf, and D. R. Shaw, *J. Immunol.*, **34**, 447 (1938).
8. Flosdorf, E. W., and S. Mudd, *J. Immunol.*, **34**, 469 (1938).
9. Harris, D. L., and L. F. Shackell, *Am. J. Pub. Health*, **7**, 52 (1911).
10. Rivers, T. M., and S. M. Ward, *J. Exptl. Med.*, **62**, 549 (1935).
11. Siedentopf, H. A., and R. G. Green, *J. Infectious Diseases*, **71**, 253 (1942).
12. Siedentopf, H. A., *U. S. Pat.*, 2,380,339.
13. Hoffstadt, R. E., and H. B. Tripi, *J. Infectious Diseases*, **78**, 183 (1946).
14. Munce, T. W., and J. Reichel, *Am. J. Vet. Research*, **4**, 270 (1943).
15. Libby, R. L., *Trans. N. Y. Acad. Sci.*, **9**, 248 (1947).
16. Wooley, J. G., *Pub. Health Repts.*, **54**, 1077 (1939).
17. Altmann, R., *Die Elementarorganismen*.   Veit, Leipzig, 1894, p. 27.
18. Goodspeed, T. H., and F. M. Uber, *Proc. Natl. Acad. Sci. U. S.*, **20**, 495 (1934).
19. Schade, A. L., and L. Caroline, *J. Bact.*, **46**, 463 (1943).
20. Teague, D. M., H. Galbraith, F. C. Hummel, H. H. Williams, and I. G. Macy, *J. Lab. Clin. Med.*, **28**, 343 (1942).
21. Johnston, C. G., *Surgery*, **3**, 875 (1938).
22. Farr, L. E., and A. Hiller, *Federation Proc.*, **II, 5**, 133 (1946).
23. Dounce, A., and J. W. Howland, *Science*, **97**, 21 (1943).
24. Flosdorf, E. W., and S. Mudd, *J. Immunol.*, **29**, 389 (1935).
25. Flosdorf, E. W., L. W. Hull, and S. Mudd, *J. Immunol.*, **50**, 21 (1945).
26. Hammer, B. W., *J. Med. Research*, 24, 527 (1911).
27. Rogers, L. A., *J. Infectious Diseases*, 14, 100 (1914).
28. Swift, H. F., *J. Exptl. Med.*, **33**, 69 (1921).

29. Siler, J. F., and the Laboratory Staff of the Army Medical School. *Am. J. Pub. Health*, **26**, 219 (1936).

30. Welch, H., E. K. Borman, and F. L. Mickle, *Am. J. Pub. Health*, **29**, 35 (1939).

31. Appleman, M. D., and O. H. Sears, *J. Bact.*, **52**, 209 (1940).

32. Seibert, F. B., and E. H. DuFour, *Am. Rev. Tuberc.*, **41**, 471 (1940).

33. Langner, P. H., Jr., and J. S. Forrester, *J. Immunol.*, **37**, 133 (1939).

34. Forrester, J. S., and P. H. Langner, Jr., *J. Immunol.*, **37**, 141 (1939).

35. Langner, P. H., Jr., and R. A. Kern, *J. Allergy*, **10**, 1 (1938).

36. Casals, J., *Science*, **97**, 337 (1943).

37. Mudd, S., *Science*, **105**, 306 (1947).

38. Hetherington, D. C., *Proc. Soc. Exptl. Biol. Med.*, **57**, 196 (1944).

39. Souter, A. W., and R. Kark, *Am. J. Med. Sci.*, **200**, 603 (1940).

40. Flosdorf, E. W., *Food Inds.*, **17** (Jan., 1945).

41. Flosdorf, E. W., *Meat Magazine* (April, 1945).

42. Gersh, I., *Anat. Record*, **53**, 309 (1932).

43. Bensley, R. R., and I. Gersh, *Anat. Record*, **57**, 205, 217, 369 (1933). R. R. Bensley, *ibid.*, **58**, 1 (1933).   I. Gersh and E. J. Stieglitz, *ibid.*, **58**, 349 (1934).   R. R. Bensley and N. L. Hoerr, *ibid.*, **60**, 251, 449 (1934).   I. Gersh and A. Tarr, *ibid.*, **63**, 231 (1935).

44. Goodspeed, T. H., F. M. Uber, and P. Avery, *Univ. Calif. Pubs. Bot.*, **18**, 33 (1935).

45. Hoerr, N. L., *Anat. Record*, **65**, 293 (1936).

46. Packer, D. M., and G. H. Scott, *J. Tech. Methods*, **22**, 85 (1942).   See also D. Glick, *Techniques of Histo- and Cytochemistry*, Interscience, New York, 1949, pp. 3–9.

47. Scott, G. H., personal communication.   See also N. L. Hoerr and G. H. Scott in *Medical Physics*, Glasser, ed., Yearbook Publishers, Chicago, 1944, pp. 466–468.

48. Stokes, F. J., Machine Co., Philadelphia 20, Pa.

BIOELECTRIC MEASUREMENTS

HOWARD J. CURTIS, *Vanderbilt University*

A. INTRODUCTION

The problem of the electrical characteristics of cells and tissues has been of interest to biologists for a great many years. Indeed it can almost be said that electricity was first discovered as a biological phenomenon. For many years biological indicators were used to detect the presence of electricity, and the strength of a voltaic pile

was tested by the strength of the reaction it could evoke from a frog's muscle. A heated controversy arose between Galvani and Volta over the question of whether "animal electricity" was different from "mineral electricity," and whether the ability of an animal to generate electricity was one of the fundamental characteristics of life. This controversy stimulated a great deal of interest in the electrical characteristics of cells and tissues, which persists to the present day.

From time to time it has been postulated that electrical forces or electric currents are responsible for the function of almost every organ of the body. These postulates have been not without some foundation, for it was soon found that electric currents were generated not only by nerve and muscle, but if careful measurements were made electric currents could be detected around practically every organ in the body.

The uses of electricity in medical therapy have not been neglected, and at one time or another it is probable that electric currents have been advocated as a remedy for practically every disease that afflicts mankind. It was not many years ago when a doctor's office was not complete without a static machine for the administration of electric treatments, and as recently as 1900 books were being published on the best methods of applying electric treatments for the cure of practically every disease. These were sincere efforts by the medical profession to find some use for this new discovery. However, medical quacks found this a most profitable field, and electric belts or similar appliances are probably still for sale in some parts of this country.

This situation could not have prevailed if there had been no element of truth in these theories and suppositions. Certain of the electric potentials found to exist in and around cells and tissues are now known to be of primary importance for the function of the cell, others are by-products of cellular function, and a great many others are measurement artifacts. As a result of a number of systematic investigations, these phenomena are now reasonably well understood, at least from a measurement point of view. There is no real excuse now for investigators to commit gross errors of measurement or interpretation as has occurred so often in the past in this field. Measuring equipment can be obtained that is accurate and inexpensive, and the sources of error are now well understood.

This chapter will attempt to outline some of the potentialities of this method and to point out some of the commonest sources of error.

The field of electrical measurements in biology can be divided
into two general parts;   potential measurements and impedance
measurements.   In studying the former one attempts to measure
the electric potentials generated by biological cells and from these
deduce something of the functional characteristics of its cells.   In
the latter, one measures the way in which an externally applied cur-
rent flows in and around cells in an attempt to learn more about the
structural characteristics of the cells and tissues.   Before consider-
ing each of these fields separately, it will be profitable to consider
certain features that both fields have in common.

1. Structure of Cells and Tissues

From an electrical point of view, a living cell consists of an
ionized salt solution called protoplasm, which is a good conductor of
electricity, surrounded by a cell membrane, which is a very poor con-
ductor.   The cell is usually immersed in a salt solution, which is a
good conductor.   Because of the nonconducting cell membrane and
the conducting solution around the cell, it is very difficult to make a
current flow through the cell.   The path of least resistance is around
the cell.   It is this short-circuiting effect of the intercellular fluid
that makes these measurements so difficult.

A tissue is an aggregation of similarly specialized cells united for
the performance of a particular function.   From our present point
of view each cell is a separate entity and is in no way dependent upon
its neighbors.   The various cells of a tissue are not all the same size
by any means, and the shapes may vary considerably.   Therefore,
except in special cases, we would not expect any special electrical
characteristics from a tissue that we would not also expect from a
miscellaneous collection of cells of similar size and shape.

2. Equivalent Circuits

Whenever an electric current is conducted from one point to
another, in any sort of medium, it should be possible to construct
an equivalent circuit, that is to say, represent the circuit in terms of
the basic electrical components.   Failure to do this in the past has
resulted in a great deal of erroneous thinking relative to electrical
problems in biology, and many apparently puzzling phenomena ap-
pear quite simple when analyzed in this elementary way.

An equivalent circuit, as the name implies, is a circuit made up of

the conventional elements of resistance, capacitance, inductance,
and electromotive force, which from all points of view is identical
with the circuit under investigation.   If one has drawn a correct
equivalent circuit, it should be possible to construct the circuit from
conventional components and demonstrate that electrically it is
identical to the unknown.   Indeed, if the experimental circuit and
its equivalent circuit were placed in identical sealed boxes with only
the two lead wires protruding from each, it would be impossible by
any electrical measurements to distinguish between them.   This
does not imply that there is only one possible equivalent circuit.
There may be a number of different equivalent circuits for any par-
ticular experimental circuit that will fulfill the above criteria, and
the one that is chosen is decided by its usefulness.

An equivalent circuit may be either drawn from the known be-
havior of the components of the experimental circuit, or constructed
from impedance and potential measurements.   In the first case, the
circuit is obtained by the laws of combination of series and parallel
circuits as demonstrated in elementary physics texts.   The latter
case is the subject of this chapter.

In general it is important to draw the equivalent circuit for one of
two reasons.   The first is to prevent errors of measurement and to
make sure that the measurements actually represent what they are
supposed to.   The second is to aid in the interpretation of the meas-
urements.

While it is usually sufficient to consider only the experimental
circuit when drawing an equivalent circuit, it should never be for-
gotten when making measurements that the measuring equipment is
an integral part of the entire circuit and when necessary its equivalent
circuit should be included in the entire equivalent circuit.   An ex-
ample will serve to illustrate the importance of the concept.   Many
investigators have tried to produce cellular changes by subjecting
cells to an electric field.   A cell suspension is usually placed in a test
tube or other suitable glass container, which is then placed between
the poles of a high d.c. potential source.   Since we are dealing with
direct currents we can ignore the membrane capacitance of the cells
and represent the cell suspension by two resistances in parallel.   The
glass walls of the test tube will act as a dielectric so there will be a
capacitance between the polepiece and the suspension on each side.
Thus the equivalent circuit can be represented by the circuit of Fig-
ure 1.   From this it will be seen that there can be no flow of direct

current in the circuit. Then as far as the resistances of the cell suspension are concerned, from Ohm's law, $E = RI = 0$. Thus as far as the cells are concerned no potential exists across them and they are subjected to no electrical influence. The field is entirely short-circuited by the salt solution. It is no wonder then that such of these experiments as have been done carefully have shown negative results. The investigators could have saved themselves the trouble of doing the experiment if they had bothered to draw an equivalent circuit.

Fig. 1. Elementary equivalent circuit for a test tube containing a cell suspension between two d.c. electrodes.

It would be difficult to overemphasize the importance of drawing a complete equivalent circuit for each measuring problem under consideration and then assigning values to as many of the components as possible. Indeed very little confidence could be placed in measurements unless this had been done before the measurements were undertaken. For these reasons, the present chapter will be developed entirely from the point of view of equivalent circuits.

B. POTENTIAL MEASUREMENTS

1. Origin of Bioelectric Potentials

It has been known for a great many years that electrical potential differences exist between the cytoplasm and the surrounding fluid of all living cells. Further it is found that for many tissues a potential difference exists between the two different sides of the tissue, for example, the inside and outside of the intestinal mucosa or inside and outside of frog skin. These potentials are not invariant quantities but depend upon the physiological state of the material and especially upon the concentration and composition of the salt solutions in contact with the cells and tissues.

A complete discussion of the origins of these potentials is beyond the scope of this book. For a more complete treatment the reader is referred to Bayliss (3), Höber (16), or MacInnes (19). However,

before any measurements in this field are attempted it will be necessary to know something of the possible sources of potentials in biological systems.

There are in general two sources one could think of as the origin of these potentials. The first is static or frictional electric charges arising at an interphase between solid and liquid, solid and air, etc. That such potentials exist in biological systems can be shown in a number of ways. For example, if a suspension of living cells is placed between two electrodes and a direct current passed through the suspension, it will be seen that the cells will move under the influence of the electric field. This means that there must be a surface charge on the cell membrane. A measurement of the speed of migration of the cell gives a measure of the electrophoretic mobility, from which it is possible to compute the surface potential under certain conditions. The importance of these potentials has been discussed by Abramson (1) and by Abramson, Moyer, and Gorin (2) (see also Chapter IX).

The second source of bioelectric potentials is quite distinct from the surface charges and is probably ionic in origin. If an electrode is placed against the cut surface of a living muscle and another placed on an uncut portion of the muscle, a potential difference will be found to exist between the two electrodes. Moreover, a current will flow between the electrodes for long periods of time. This fact shows conclusively that the potential cannot be of static origin, since if it were it would be rapidly dissipated. This potential is also probably not directly due to a chemical reaction, since the electromotive force is roughly proportional to the absolute temperature. Most chemical reactions have temperature coefficients much larger than this.

The dissociation of an electrolyte into positive and negative ions cannot of itself give rise to a potential, since the ions are always present in equal numbers. However, under certain circumstances the two kinds of ions may become somewhat separated from each other, giving rise to local potential differences. This is probably the origin of most bioelectric potentials usually measured. It remains to be considered how this separation takes place.

Suppose that the ions are free to diffuse away from their point of origin after they are formed. In general one ion will diffuse faster than the other and in this way they will become somewhat separated. For example, if a tube containing a strong sodium chloride solution is dipped into a beaker containing a weak sodium chloride solution in

such a way that the two solutions come in contact with each other, a potential will be found to exist between the two solutions as long as the difference in concentration persists. These potentials are known as liquid junction potentials. In 1889 Nernst derived an equation expressing the voltage, E, obtained from a liquid junction:

$$E = \frac{RT}{n} \frac{u - v}{u + v} \ln \frac{C_1}{C_2} \tag{1}$$

where C_1 and C_2 are the concentrations in the two solutions, u and v the mobilities of cation and anion, respectively, R the gas constant, T the absolute temperature, and n the quantity of electricity carried by one gram equivalent. It will be noted that the voltage depends on the difference in mobility of the two ions. For example, in the case of hydrochloric acid, the hydrogen ion can diffuse about five times as fast as the chloride ion. At a liquid junction both ions would tend to diffuse from the concentrated solution to the dilute, but the hydrogen going faster would make the dilute solution positive with respect to the concentrated solution by an amount given by equation (1). If the mobilities of the two ions are practically the same, as is the case with potassium chloride, the liquid junction potential practically vanishes.

Equation (1) has been found to be very nearly correct for the potential between two solutions having practically the same activity coefficient. If this condition is not fulfilled, equation (1) must be somewhat modified.

The question arises as to what part, if any, these liquid junction potentials play in the potentials commonly measured in biological systems. As we proceed with a discussion of methods of measuring these potentials it will become evident that, in any actual measuring set-up, liquid junctions play quite a prominent part in determining the measured potential. However, whether liquid junction potentials normally exist in and around normal single cells is quite doubtful. The distances involved are usually small and diffusion equilibrium could take place quite rapidly. Thus it is usually felt that except for special cases liquid junction potentials as they normally appear represent measurement artifacts.

If, however, one ion of an electrolyte is constrained so that it cannot move, while the other is free to migrate, it is clear that an appreciable potential could develop. Such a situation exists when two solutions are separated by a membrane that will allow only one of

the ions of an electrolyte to pass through. This can be created in the laboratory by separating two different solutions by a membrane having pores so small that one of the ions in one of the solutions cannot pass through—for example, a solution containing an ionized protein separated from a salt solution by a cellophane membrane. Even after equilibrium has been established, a potential difference will exist between the solutions. This has come to be known as a membrane potential, and Nernst proposed the equation:

$$E = \frac{RT}{n} \ln \frac{C_1}{C_2} \tag{2}$$

as a measure of this potential. It will be seen that the equation is identical to equation (1) for the case in which the mobility of one ion is very much larger than the other, so Nernst's view of the membrane potential would merely be a special kind of liquid junction potential. On the basis of this equation the membrane potential would amount to about 0.057 v. at room temperature for a 10 to 1 difference in concentration of the diffusible ion between the two sides of the membrane at equilibrium. Whereas this formula has been verified in a general way, it is certainly not completely correct. A number of efforts have been made to place it on a more quantitative basis, none of which has been completely successful.

Certainly the best explanation of the potential difference that exists between the inside and outside of living cells is to be found in this theory. Due to measurement difficulties, which will be discussed later, there has been very little actual verification of this. However, such measurements as have been made tend to verify the general concepts of the Nernst hypothesis, but leave many details unexplained. According to this hypothesis, the living cell contains quite a high concentration of large organic anions that cannot diffuse through the cell membrane. Further, of the common cations found in biological systems, the cell membrane is presumed to be permeable only to potassium. Under these conditions, equilibrium will be established only when the concentration of K^+ inside the cell is higher than the concentration outside. Then the tendency of the K^+ to diffuse out and equalize the K^+ concentration inside and out will be balanced by the tendency of the K^+ to diffuse in and neutralize the excess anions inside. On this basis the membrane potential would be given by equation (2), where C_1 and C_2 represent the K^+ concentrations inside and outside the cell. And if one were to plot the log-

arithm of the K^+ concentration outside the cell as a function of the
membrane potential, a straight line should be obtained. Figure 2
shows the results of a set of measurements taken on the giant nerve
cell of the squid. Here the membrane potential is plotted against
the logarithm of the K^+ concentration outside the cell. When the
K^+ concentrations inside and outside are equal, there is no membrane
potential. However, as the K^+ concentration outside the cell is
reduced, the negative potential of the cytoplasm of the cell increases
in a straight line until quite a low value of K^+ concentration is
reached, and then this relationship breaks down. Why the Nernst

Fig. 2. Membrane potential
from the squid giant axon, meas-
ured between one electrode in-
side the cell and another out-
side. The potential is measured
as a function of the relative K^+
concentration outside the cell
compared to the normal K^+
concentration of sea water (12).
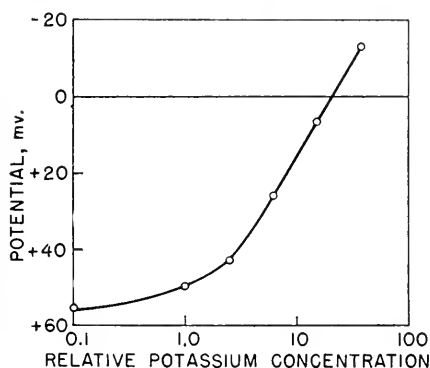
relation is not followed beyond this point is not known, but presum-
ably some of the postulates are no longer valid. Experiments such
as these give us considerable confidence that the above explanation
for cell potentials is correct in a general way.

2. Electrodes

In making potential measurements in biological systems it is al-
ways necessary to make an electrical connection between the meas-
uring device and the electrolytic solution of the biological system.
This involves a metal-to-liquid junction. Ideally one could think
of finding a metal that would show no potential difference between it-
self and an aqueous solution. This appears to be impossible and the
problem has resolved itself into one of finding an electrode that will
give constant and reproducible potentials when placed in contact
with any salt solution.

When a metal is placed in water there is a tendency for some of
the metal to go into solution. This is known as the solution pressure

of the metal. Thus if copper is placed in water some Cu^{++} will migrate from the metal into the water, leaving the metal negatively charged with respect to the water. Ions will continue to leave the metal until the negative potential developed is strong enough to prevent further escape. This potential is very unstable, and any movement of the water, addition of foreign ions, etc. would upset the equilibrium. However, if the metal is in a solution of one of its salts, for example, copper in copper sulfate, the Cu^{++} will tend to deposit on the metal, tending to increase the positive potential of the metal. Thus there would be a tendency for the Cu^{++} to leave the metal and go into solution and also a tendency to be deposited on the metal from the solution, and so an equilibrium would be established. This causes a potential between the metal and the solution that is quite stable and depends only upon the metal, the salt concentration, and the temperature. It is known as a reversible electrode since, if current flows from the copper to the solution, copper goes into solution and if the current is reversed the reverse is true quantitatively.

There have been many theories and a good deal of experimental work attempting to determine the absolute potential of an electrode. Whereas some of the theories have been quite successful, the problem is still not in a completely satisfactory state, and the reader is referred to a text on physical chemistry for a complete discussion of the basic theory of electrode potentials.

In any real system it is of course necessary to have two electrodes, and it is never possible to do more than measure the potential difference between two electrodes. For practical reasons the standard hydrogen electrode has been taken as a zero of potential, and all unknown electrode potentials referred to it. It is not customary to think of a gas as an electrode, but it can function very satisfactorily as such under proper conditions. The standard hydrogen electrode (see Glasstone, *15*) is made by placing an electrode of platinized platinum half in an atmosphere of hydrogen gas and half in a solution of sulfuric acid of such a concentration that it contains 1 g. of H^+ per liter. Hydrogen gas is bubbled through the solution until equilibrium is established. There are a number of precautions that must be taken in using this electrode, which will not be treated here, and when these are taken it makes a very good standard electrode. However, for practical reasons it is usually more convenient to use a calomel electrode for everything but very precise chemical work.

The Calomel Electrode. This type will maintain a constant and reproducible potential if only reasonable precautions are taken in preparing the electrode.   For this reason, it has been extensively used for potential measurements in biology.   It will even maintain a reasonably constant potential during current flow, which is a very stringent requirement.

There are a number of different satisfactory forms of the calomel electrode, one of which is shown in Figure 3.   A platinum wire sealed into tube A makes contact with the mercury at the bottom of the tube.   This is covered with a paste made by grinding mercurous
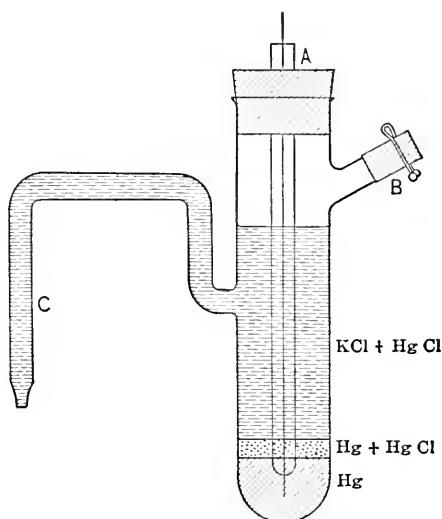
Fig. 3.   Calomel reference electrode (*14a*).

chloride with mercury together with a small amount of molar potassium chloride in a mortar.   This is covered with a molar solution of potassium chloride saturated by prolonged shaking with mercurous chloride.   Contact is made to the liquid under investigation by means of tube C.   The potential of this electrode is $+0.282$ v. at $25°$ referred to the standard hydrogen electrode as zero.   The plus sign indicates that the electrode is positive to the solution.

If it is necessary to maintain a constant potential in the face of a current flow, the success achieved will be roughly proportional to the area of the mercury in contact with the solution.   Electrodes as large as 8 cm. in diameter have been used.

Silver–Silver Chloride Electrodes. The calomel electrode is rather bulky and inconvenient, so silver electrodes in the presence of a saturated solution of silver chloride are often used. Silver is usually electroplated on a gold electrode of any convenient shape to insure having a pure silver surface. This electrode is then made the anode in an electroplating cell using a pure solution of sodium or potassium chloride. Current is passed slowly over a considerable period of time until a thick adherent coating of silver chloride covers the entire electrode. The electrode is then ready for use, and it may be placed directly in the liquid in question or against the tissue, or may be placed in a potassium chloride solution and connected to the solution in question by means of a salt bridge. The electrode should never be allowed to dry out, and if accurate absolute values are desired the electrodes should never be exposed to the light. It will be clear that in use the silver electrode will be surrounded by a saturated solution of silver chloride, since the silver chloride coat will very slowly dissolve and it takes only a minute amount to form a saturated solution. This then fulfills all the requirements for a reversible and reproducible electrode.

It is possible to make these electrodes so that they will give very accurate results. However, the requirements that must be met are more stringent than in the case of calomel electrodes. Therefore it is not recommended that these electrodes be used as reference electrodes except in special cases. Their greatest usefulness is as working electrodes.

Working Electrodes. In many biological investigations there is no interest in the absolute values of electrode potentials, but the electrodes are used in pairs of identical electrodes so the electrode potentials, being equal and opposite, cancel out. It is only necessary then to make sure that the two electrodes are identical. It is customary to make electrodes in pairs some days before they are needed. They are then stored in approximately the same solution in which they will be used, and short-circuited. Usually slight differences of potential are equalized in this way. These electrodes will remain quite constant even during the passage of current. However, the larger the current to be passed, the larger the surface area of the electrode should be.

A very convenient form for these electrodes is shown in Figure 4. The electrode itself is in the form of a helix of silver wire (silver plating is unnecessary here) immersed in a tube completely filled with a

solution of potassium chloride isotonic with the solutions to be measured. The tube is sealed with a rubber stopper and contact is made through a camel's hair brush or cotton wick. Chloride is plated on after the electrode is finished. These should always be made and used in pairs. They are very convenient to use and may be stored for months or even years and used on a moment's notice.
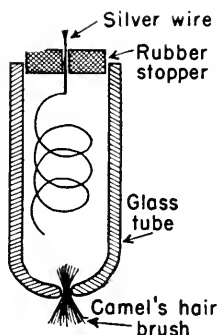
Fig. 4. Convenient working electrode for biological investigations.
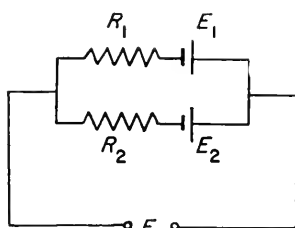
Fig. 5. Equivalent circuit for potential measurements, showing magnitude of error when leakage is present.

3. Salt Bridges

It is usually desirable to isolate the electrode itself from the solution being measured, and this is done by means of a salt bridge. It is merely a means of making electrical contact between two solutions without introducing electrodes and their accompanying complications. It usually consists merely of an inverted U tube filled with molar potassium chloride solution, with one end dipping in each of the solutions. It will be seen from equation (1) that under these circumstances there will be a liquid junction potential between the solution of the salt bridge and the other solutions. However, it will be seen that the potential is proportional to the *difference* in mobility between anion and cation. The mobility of K^+ and Cl^- are almost identical, so a potassium chloride solution can add nothing to a liquid junction potential. Thus if two solutions are joined electrically by a potassium chloride salt bridge, they will have no more of a potential between them than they would have if it were possible to place the two solutions in direct contact.

It is sometimes convenient to make these bridges out of a potassium chloride solution in an agar gel. This is permissible except where very exact measurements are required. There may be small spurious potentials, probably introduced by impurities in the agar, which cannot be predicted and in general will be different at the two ends of the bridge.

4. Insulation

When making potential measurements in aqueous systems one must exercise extreme care to make sure that there are no leakage currents. Such currents often give rise to measured potentials much larger than the potentials being measured. Many apparently startling discoveries have been found to be due to this cause. Proper insulation is not difficult, but certain precautions must be taken from the start and constantly checked.

In making such measurements it is necessary to so arrange things that the resistance through the desired path is low relative to that through any other path. In most cases the equivalent circuit for such a system can be reduced to that shown in Figure 5. Here the circuit under investigation is represented by R_1E_1 and the object of the measurement is to measure the magnitude of E_1. In general there will be some leakage between the two electrodes. The leakage resistance, R_2, will in general have a potential associated with it, represented here by E_2. The potential measured at the measuring instrument, E, will be:

$$E = E_1 - \frac{R_1(E_1 - E_2)}{R_1 + R_2} \tag{3}$$

If R_2 is very much larger than R_1, $E = E_1$ and the measured value will be correct. However, if R_2 is the same size as R_1, as may well be the case in an actual experiment:

$$E = (E_1 + E_2)/2$$

and the spurious potential will play as large a part as the potential being measured. Under such conditions the measured potential would be meaningless.

It is not always easy to make sure that the spurious potentials are not significantly influencing the measured potential. A common practice is to short-circuit the measured potential with a salt bridge

from one electrode to the other, and assume that any potential measured then is due to electrode differences or to spurious potentials, and then to use this value as a zero point for the actual measurement. The fallacy of this procedure can be seen by glancing at equation (3). The salt bridge merely changes the value of R_1 and eliminates part of E_1, whereas the important consideration is the relation between the actual value of R_1 and R_2 of which this short-circuiting procedure may give no indication.

It is impossible to outline any procedure that can be used to insure elimination of these errors in all cases. The investigator must be fully aware of the possible sources of error, draw an equivalent circuit for the particular set-up being used, and show from this that all sources of error have been removed. Until this is done, it is impossible to have any faith whatever in the measurements.

5. Insulating Materials

When using insulating materials for biological work it is usually necessary to develop criteria of excellence different from those used in the physical sciences. In the latter, volume resistivity is usually rather important, since surface conductance can be kept small by keeping the insulator clean and dry. However, this is very difficult if not impossible in biological work, so volume resistivity is of very little importance. Any insulator impervious to water will in general have a high enough volume resistivity to be acceptable for biological work. Glass and quartz, for example, have a high volume resistivity but are readily wet by water and so are bad from the biological point of view, especially if they are clean. Perhaps the best substance known is paraffin, which has both a very high surface and volume resistivity and is not wet by water. However, it is rather inconvenient to use in many cases. Paraffin-coated glass has often been used, but extreme care must be used, since in the course of time a film of water will work between the glass and the paraffin and cause trouble while the coating appears to be in good condition.

Glass coated with petroleum jelly or other heavy mineral oil is very good, although the coating must be renewed from time to time. This can be done merely by wiping the glass with a cloth saturated with petroleum jelly. Some of the plastics such as lucite or polystyrene are quite good by themselves, and very good when coated with petroleum jelly.

C. EQUIPMENT FOR POTENTIAL MEASUREMENTS

The problem of equipment for measuring these potentials is as old as the science of electrobiology, and still continues to be a problem. Each time there has been an improvement in electrical measuring equipment in physics or engineering, someone has been quick to apply it to problems of potential measurements in biology. In some cases new types of electrical measuring equipment have been devised by biologists to meet the stringent demands of particular biological problems. For example, Einthoven developed the string galvanometer solely for the purpose of measuring the electrical potential changes associated with the heart beat.

Prior to the development of the radio tube, a number of instruments were used for the measurement of biological potentials. Among these can be mentioned the quadrant electrometer, the string galvanometer, the moving-coil galvanometer, and others. All have been important in the development of electrophysiology, but only the moving-coil galvanometer retains any degree of importance. For a general treatment of these older instruments, the reader is referred to Bayliss (3).

In general, it is now customary first to amplify the potential being measured by means of an electronic amplifier and then apply the amplified potential to some electrical measuring instrument. The apparatus can then be discussed in two parts.

1. Electronic Amplifiers

A complete discussion of electronic tubes is far beyond the scope of this chapter. However, these tubes are extremely useful devices in biological research, and the biologist can use them intelligently without knowing the complete theory behind them in much the same way that he uses a microscope without knowing all of the theory of optics that went into the construction of the instrument. The biologist is advised against taking a course in communications engineering since the material presented there does not fulfill his needs. Instead, he is referred to a book by Muller, Garman, and Droz (20), which contains the elements of experimental electronics as they are needed by the biologist.

With a relatively elementary knowledge of electronic tubes and circuits it is possible for the biologist to construct and use this equipment from published circuit diagrams. As he becomes more profi-

cient in their use he can make changes to fit his particular needs, and later will not find it difficult to design his own equipment.   With this in mind, several electronic circuits will be reproduced here without elaborate explanation.

Electronic amplifiers are used in measuring bioelectric potentials for two reasons: first because of their very high input resistance, and second because of their speed of response.   When a measuring instrument or amplifier is connected to the electrodes in an aqueous system, the instrument becomes an integral part of the equivalent circuit and may significantly alter the potential distribution and thus cause considerable error.   It will be realized from the above discussion of measurement errors that errors can be avoided only if the resistance of the measuring instrument is very much higher than that of the circuit being measured.   The resistance of many biological circuits can be thousands and even millions of ohms, so in general the input resistance of a satisfactory measuring instrument must be many megohms.   Electronic amplifiers and electrometers are the only instruments satisfying this condition.   The latter is extremely difficult to set up with the requisite sensitivity, and even so is a very sluggish instrument, difficult to use and extremely sensitive to moisture.   On the other hand, the electronic amplifier can be constructed with almost any desired input resistance and sensitivity, and can be made to actuate almost any desired recording instrument.

Amplifiers may be put in three general classes for bioelectric potential measurements: (1) Very high resistance (electrometer type), in which the primary consideration is a high input resistance; (2) general purpose, direct-coupled amplifiers;   (3) capacitor-coupled amplifiers for measuring rapid changes in potential.   Each of these will be briefly considered.

2. Electrometer Amplifiers

An electronic tube has often been described as a valve in which the flow of electrons between cathode and plate is regulated by the potential on the grid.   According to this view, there would be no flow of current in the grid circuit at all, and indeed this is true to a first approximation.   However, some of the electrons in the tube adhere to the grid, causing some grid current, and this can become quite large if the grid becomes positive with respect to the cathode.   Also, there are always a few positive ions present in the tube due to electron bombardment of the residual gas of the tube, and these will be at-

tracted to the grid provided it is negative. Thus the grid current will be positive or negative depending on the magnitude and sign of the grid potential. It then follows that at some grid potential the grid current will be zero and the grid resistance (input resistance) will be infinite at this potential. This cannot be fully achieved, but some very close approximations can be made.

Actually, of course, there are many causes of grid current, many of which can be minimized by careful tube design and construction. Such tubes are manufactured, and are known as electrometer tubes. In such tubes all other characteristics are sacrificed in favor of a low grid current.

A large number of circuits have been published for use with these electrometer tubes, each desirable for a particular application. However, there has been a large demand for such circuits to be used in connection with glass electrodes. A number of commercial companies now manufacture these glass electrode amplifiers, and they function very satisfactorily. It would therefore be inadvisable for a biologist to attempt to build one of these amplifiers, since their design and construction are quite critical. The commercial glass electrode amplifier can be adapted readily to almost any use requiring a very high resistance input. For special applications it may be necessary to construct such an amplifier, and for this the reader is referred to articles by Turner (*22*), DuBridge and Brown (*13*), Penick (*21*), and Bearden (*4*).

3. Direct Current Amplifiers

It is seldom necessary, in biological work, to resort to the use of electrometer type amplifiers, and a general purpose amplifier will usually suffice. With almost no precautions such amplifiers draw an input current of 10^{-7} amp. or less and with reasonable precautions will go down to 10^{-9} amp. or less. If it is necessary to go below 10^{-10} amp., an electrometer circuit should be employed.

A curve of grid current versus grid voltage is shown for a typical triode in Figure 6. It will be seen that if the grid voltage is always less than -1 v., the grid current will never be more than about 10^{-9} for this particular tube. However, if the grid voltage becomes more positive than about -0.5 v., the grid current may be relatively large. This curve shows the fallacy of the statement one often hears that the grid current will be negligible as long as the grid voltage is negative. From the curve of Figure 6 it will be seen that at about -0.75

v. the grid current is zero. Thus, if provision is made for running
the tube always exactly at this voltage, the grid current can be kept
extremely low.   This principle forms the basis of several instruments
designed for biological work in which it is desirable to keep the grid
current low.   However, these circuits are somewhat awkward to use
and are not recommended.

It should be emphasized that the curve of Figure 6 is for one par-
ticular triode, operating at one particular plate voltage. Other
tubes show curves of the same form but the magnitude of current and
voltages may be quite different.   If there is any question as to how
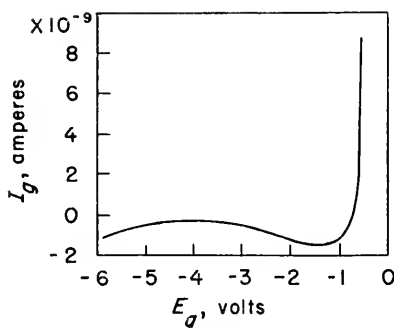
Fig. 6.   Grid current *vs.* grid voltage for a typical triode.

low the grid current is in any particular amplifier, it should be meas-
ured directly.   This can be done conveniently by placing a high re-
sistance (100 megohms or more) of known value in series with the grid
and noting the plate current.   Then the resistance is replaced by a
low resistance potentiometer and a potential applied that will give
the same plate current.   The grid potential must have been the same
in both cases, which means that the potential drop across the resistor
is known.   From this, and knowing the value of the resistor, the
grid current can be computed from Ohm's law.   If a very low grid
current is necessary, it is desirable to plot out a grid current versus
grid voltage curve for the particular tube being used and the par-
ticular way in which it is used.

One of the most troublesome features of direct current amplifiers
is drift.   It is practically impossible to eliminate drift completely,
but a few simple precautions will enormously reduce it.   The first
thing to do is to use a balanced amplifier throughout, so the drift of

one tube is balanced by the drift of another. The second is to use constant voltages, especially in the first stages. This can be done moderately well by the use of batteries, but even these drift some. A much better way is by the use of an electronically regulated power supply (see Sect. C6). The third precaution is the selection of tubes. The best tubes for this purpose are of course the electrometer tubes. However, they are expensive and have characteristics that are not always suitable for particular amplifiers. The author has found that it is reasonably satisfactory to use tubes such as the type 12SJ7, run as triodes in the first stages and as pentodes in subsequent stages. It is best to purchase a number of these tubes all at the same time and from the same dealer, run them at a moderate plate current for
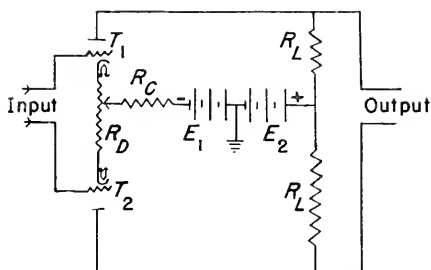
Fig. 7.   Diagram showing common mode degeneration.

100–300 hours, and then carefully match them in pairs. Put the best pair in the first stage, the next best pair in the second stage, etc.

There are a number of other things that can be done to minimize drift. The plate voltages and currents should be kept as low as possible; voltages as low as 20 v. are not unreasonable. Filament current on individual tubes can be carefully adjusted. Precautions can be taken to keep the temperature changes the same on both tubes of a pair. Wire-wound resistors can be used throughout. Indeed, there is almost no end to possible refinements, most of which are unnecessary for biological work.

Another troublesome feature of amplifiers is pick-up. Very elaborate precautions have been taken in the past to prevent induced currents from being amplified along with the potentials under study. These precautions usually consisted of extensive shielding, and even elaborately shielded rooms. However, the use of amplifiers with common mode degeneration has made the use of such shielding com-

pletely unnecessary.   The principle of the balanced amplifier with
common mode degeneration can be illustrated best by means of a
diagram such as shown in Figure 7.   It will be seen that the cathode
current of both tubes flows through the resistor R_C causing a potential
drop that tends to make the cathode positive with respect to ground.
The value of R_C should be such that the potential drop across it is
greater than the value of E_1 by the desired grid voltage.   This means
that the grids will be at ground potential.   The variable resistance,
R_D, should be very low compared to R_C and is used merely for balanc-
ing the two tubes.   Let us assume that a potential is applied across
the input terminals such that the grid of T_1 is made negative with
respect to the grid of T_2.   Then the current through T_1 will decrease
and that through T_2 will increase by exactly the same amount.   The
current through R_C is unchanged and therefore the potential of the
cathodes will remain unchanged.   The potential across the output
will be an amplified version of the input potential.   However, sup-
pose now that the potential of both input terminals is raised relative
to ground.   The current through both tubes will tend to increase.
This will tend to increase the current through R_C and consequently
the potential drop across R_C and thus raise the potential of the cath-
odes.   If R_C is large enough, the cathodes will increase practically as
much as the grids, leaving the effective grid potential practically un-
changed.   This would mean that the plate current would remain
almost unchanged and the output would remain unaffected.   Thus a
signal common to both grids (common mode) is degenerated prac-
tically to zero, while signals applied between the two grids (differen-
tial signal) is amplified.

Set-ups can always be arranged so that stray induced currents
are applied common mode but the potential in question is applied dif-
ferentially.   With careful adjustment the ratio between the effect
produced by a common mode and a differential signal can be made
as high as 25,000 to 1.   However, a ratio of a few thousand to one is
usually ample for most situations.   This simple principle of common
mode degeneration is of inestimable value in measuring bioelectric
potentials.

One of the difficult features of direct current amplifiers has been
the problem of coupling the plate of one tube, at a potential of a
hundred volts or so, to the grid of the tube in the next stage, which is
usually run at ground potential.   Many ingenious devices have been
worked out for doing this, most of which are unsatisfactory for one

reason or another. By far the best method, in the author's experience, is to allow the potentials to cascade. If the plates of the first stage are at a potential of $+25$ v. (with a plate supply voltage of $+50$ v.), the grids of the second stage should be connected directly to them and run at $+25$ v., with the cathodes of the second stage slightly higher, say, $+26$ v. Such an arrangement is illustrated in Figure 8. The cathode resistor (R_C) in the second stage should be variable so that it can be varied until the proper potential on the cathode is secured. The potentials can be checked by means of a commercial vacuum tube voltmeter, which is an indispensable aid in the construction of electronic equipment. When this amplifier has been balanced, the deflection of the meter, M, will be proportional to the differential input voltage.
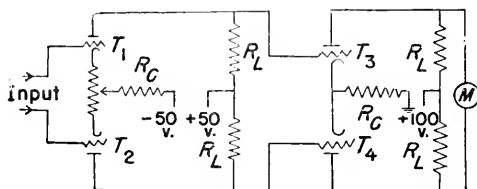
Fig. 8.   Diagram showing cascade method of interstage coupling.

The cascade method of interstage coupling can be used for almost any number of stages. It is only necessary to have a voltage increase of 30 to 60 v. per stage so a four stage amplifier can be run from a single 300 v. power supply. However, in using an amplifier with low plate voltages, care must be taken not to allow the signal to overload the last stages.

4. Capacitor-Coupled Amplifiers

If it is only necessary to measure relatively rapid changes in potential, such as the action potential of nerve or muscle, capacitance coupling between stages is quite satisfactory, and indeed is much more convenient to use than direct coupling. Direct coupling should never be used unless it is absolutely necessary. The balanced amplifier with common mode degeneration should always be used even for capacitor-coupled amplifiers. The advantages gained are worth many times the slight extra cost.

Almost the only special precaution necessary for a capacitor-coupled amplifier is to be sure to use good capacitors for interstage

coupling. Either oil or mica capacitors should be used, and only the best grades of each. This is necessary because a very slight leak from the high potential plate of one tube through the coupling capacitor can cause a large change in the grid potential of the next tube.

The size of the capacitor depends on the frequency pass required. If a steady potential is suddenly applied to the input of a capacitance-coupled amplifier, the potential on the grid of the second stage, instead of rising suddenly to a new value and staying there, will suddenly rise but then decay exponentially to its former value. The rate of decay will be a function of the size of the capacitor and the value of the grid resistance. Its time constant, t, is given by the relation:

$$t = RC \tag{4}$$

where R is the value of the grid resistor in ohms and C the value of the coupling capacitor in farads. The time will then be given in seconds, and represents the time required for the potential to fall to $1/e$ of its initial value ($e = 2.718$). The potential on the grid of the third stage will fall correspondingly faster, and so on for succeeding stages. From these facts it should be possible to determine how much error will be committed in any individual problem by using a capacitance-coupled amplifier.

From this discussion one might infer that it is best to use as large a capacitor as possible. This is true from the standpoint of accuracy, but is certainly not true for ease of operation. The larger the capacitors the longer one has to wait for equilibrium to be established in the amplifier, and, with 4 μf. capacitors in a four stage amplifier, this can be very annoying. For the recording of most action potentials from nerve or muscle, 0.1 μf. capacitors are large enough.

5. Limitations of Amplifiers

It might be inferred from the above discussion that there is no lower limit to the potentials that can be measured with electronic amplifiers. But such is certainly not the case. The ultimate limit of sensitivity is imposed by "tube noise." This can be defined as random spurious changes in plate current of a tube, and may be due to a variety of causes. The chief ones are irregularities of electron emission from the cathode and changes in resistance of resistors in the circuit. There are a number of things that can be done to reduce this artifact. The first is to use only wire-wound resistors in the first

stage of the amplifier. The next is to keep the plate potential and plate current as low as possible. If carbon potentiometers are used in the first stages they should be kept very clean with an organic solvent, while wire-wound potentiometers should be oiled on the rubbing surfaces with petroleum jelly. The tubes should be insulated from noise and vibration, although this precaution is necessary only in extreme cases with modern tubes.

When these precautions are taken, and using a type 12SJ7 tube as a triode, the noise in the plate circuit of the tube is roughly of the same magnitude as is caused by an applied potential in the grid circuit of 3 μv. Thus we speak of the equivalent noise level as being 3 μv. It is obvious from this that if a potential of less than 3 μv. were applied to the grid of the tube, it would not be possible to detect it, since it would be lost in the noise of the tube. If additional stages of amplification are added, they amplify the noise as much as the signal, so nothing is gained. The situation is analogous to that of empty magnification with a microscope—where an object possesses structure too small to be resolved by the objective, no amount of magnification will help.

This limitation can be partially overcome in special cases. If alternating current of a special frequency is amplified by an amplifier tuned to that particular frequency, only that fraction of the noise having that particular frequency will come through and the equivalent noise level can be drastically reduced. Again, if only direct currents need be amplified, it is convenient to put a capacitor across the output to short-circuit the high frequency components of the noise.

Another troublesome feature of direct current amplifiers is drift. As indicated above, this can be minimized by a number of precautions but never eliminated. A careful selection of tubes in the first stage of a balanced amplifier is probably the most important single factor. It will usually help, in a balanced amplifier, if there is a separate control on the filament current of each tube, and a value for each tube can usually be found by trial that will very considerably reduce the drift. It is not difficult to construct a d.c. amplifier in which drift is not troublesome provided potentials of a few hundred microvolts are to be measured. If the potentials are only slightly larger than the noise level, it is quite difficult to construct an amplifier in which drift is not very troublesome.

It should be realized that there are certain situations in which a galvanometer can be used more advantageously than an amplifier.

This is true when the resistance of the source is quite low so an appreciable current may be drawn from it. This is true in such devices as the barrier layer photocell or the thermocouple. Here 1 μv. will produce a current of about 10^{-10} amp. through a high resistance gal-
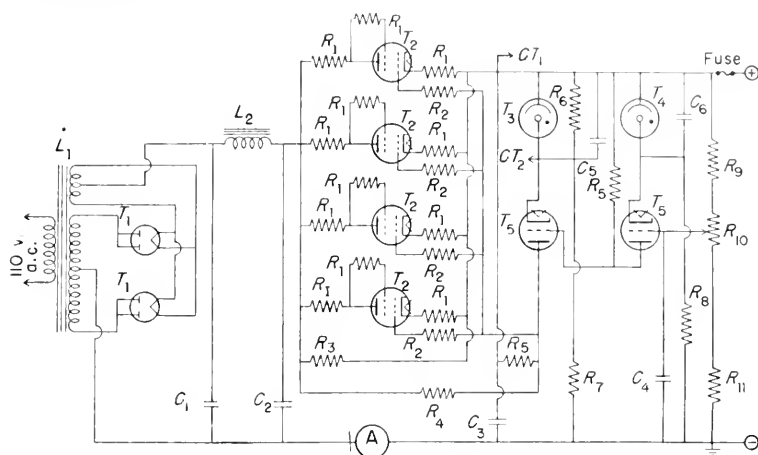
Fig. 9. Electronically regulated power supply (see text p. 258). The components are:

L_1	Power transformer: pri = 110 v., 60 cycles/sec.; sec_1 = 5 v. at 6 amp.; sec_2 = 860 v. at 450 amp.	R_2	500 Ω, 1 w.
		R_3	Power resistor (use only if current is over 250 ma.) = $\dfrac{150}{\text{load (amp.)} - 0.25}$
L_2	450 ma. filter choke		
T_1	5U4	R_4	5 megΩ, 1 w.
T_2	6L6	R_5	0.5 megΩ, 1 w.
T_3	VR-105-30	R_6	0.1 megΩ, 0.5 w.
T_4	VR-150-30	R_7	15,000 Ω, 10 w.
T_5	6SF5	R_8	10,000 Ω, 10w.
C_1	10μf. 1000 v., Dykanol	R_9	0.15 megΩ, 2 w.
C_2	8 μf., 600 v., electrolytic	R_{10}	0.10 megΩ, 2 w. potential
C_3	8 μf., 400 v., electrolytic	R_{11}	0.05 megΩ, 2 w.
C_4	4 μf., paper	A	0–500 milliammeter
C_5	0.25 μf., 150 v.	CT_1	connect to center tap of 6L6 filament transformer
C_6	0.25 μf., 200 v.	CT_2	connect to center tap of 6SF5 filament transformer
R_1	50 Ω, 1 w.		

vanometer, which will be easily registered if the galvanometer is quite sensitive. As has been indicated, 1 μv. is well below the limit of resolution of the best amplifier. Further, one does not need to worry

about drift in a galvanometer.   On the other hand, for high resistance sources a vacuum tube amplifier has a tremendous advantage over the galvanometer even for d.c. measurements.

6. Power Supplies

Batteries are certainly the easiest satisfactory means of supplying power for an amplifier.   However, they are expensive, awkward, and are continually running down, so that battery-operated amplifiers require constant attention.   The usual type of power supply is not stable enough for most work, but electronically regulated power supplies have proved to be eminently satisfactory.   One such power supply is shown in Figure 9 (p. 257).   Here the four power tubes (6L6) in parallel act as a variable resistance in series with the filtered power supply.   Any changes that take place in the output voltage are amplified and applied to the grids of these power tubes to change their effective resistance by the amount necessary to bring the output potential back to normal.   This power supply will deliver up to 450 ma. at 300 v.

The author has recently built a high gain d.c. amplifier in which a single regulated power supply furnished current for both plates and filaments of all tubes.   The tubes had filaments requiring 12 v. and 0.15 amp., and all filaments were run in series.   Operation was as stable as if the first stages were run entirely on new batteries.

7. Recording Equipment

A wide variety of different instruments has been developed for recording bioelectric potentials.   Most of the present day instruments were originally developed either for the purpose of obtaining an instrument with a faster response, or a greater sensitivity than was previously possible.   Since vacuum tube amplifiers are now so satisfactory, sensitivity is no longer a major consideration except in special cases.   So speed of response and convenience have become the major factors in the choice of instruments.

For either visual observation or photographic recording where speed of response is a consideration, the instrument of choice is a cathode ray oscillograph.   For practical purposes the instrument has no inertia whatever and recordings thus represent a true pattern of the potential changes applied to the instrument.   There are a number of very satisfactory instruments on the market, complete with all

power supplies and some auxiliary amplifiers.   In general these are
useful only for measuring potential changes, since the amplifiers are
usually capacitance-coupled.   It is possible to modify these instru-
ments, or build one from parts, in which the plates are coupled di-
rectly to the amplifier output so both direct currents and potential
changes may be measured.   A description of these instruments and
how they can be constructed is beyond the scope of this book, and
the reader is referred to such books as the one by Muller, Garman,
and Droz (20) for further details.

It is often convenient to record potentials on moving paper, either
photographically or by other means.   Of these, the photographic
methods are by far the most accurate but not the most convenient.
Recording oscillographs are available which use a mirror galvanometer
that will accurately reproduce an alternating current up to 5000 cycles
per second.   They give a linear displacement on the moving paper,
and may have a total displacement of 12 cm.   However, they all have
the disadvantages of a photographic method.

The direct-recording oscillographs do so either by a moving pen
(ink writers) or some other marking device such as a heated stylus
marking on paraffin paper.   These are definitely limited in fre-
quency response, usually being able to record frequencies only up to
about 100 cycles per second.   Also, they record over the arc of a
circle and for a width of only about a centimeter.   In general they
are more satisfactory for qualitative than for quantitative work.
However, they are very convenient to use, and this feature offsets
their disadvantages for many applications.

For very slow changes, or for measuring direct currents, an ordi-
nary milliammeter is satisfactory for most purposes.   A permanent
record may be obtained by using a recording milliammeter.

D. BIOELECTRIC POTENTIALS

1. Membrane and Action Potentials in Nerve and Muscle

As indicated in a previous section, an electrical potential exists
between the inside and outside of living cells that gives a clue to the
function of the cell, and for this reason has been extensively studied
in the past.   It is known as the membrane potential and can be
measured either directly or indirectly.   In the direct method an elec-
trode is inserted inside the cell and a measurement of the potential is

made between this inside electrode and one on the outside surface of
the cell directly opposite it (see 12).   Here there can be no question
but that the true membrane potential is being measured, provided
elementary precautions are taken about insulation, input resistance
to the amplifier, etc., as discussed above.   However, this method suf-
fers from the disadvantage that the manipulative procedure of getting
an electrode inside a cell is very difficult, and indeed has been ac-
complished for only two cells, the giant nerve fiber of the squid and
the large single plant cell *Valonia*.

This measurement can be made indirectly in the case of certain
cells as the injury potential.   Consider, for example, a nerve cell
that is a cylindrical cell in which the length is very much greater than
the diameter.   It has been found that one end of a nerve fiber can
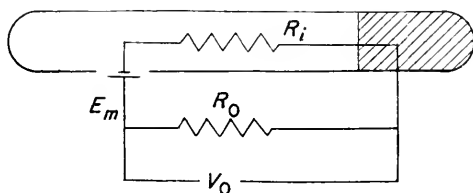
Fig. 10.   Equivalent circuit showing relation between membrane
potential E_m and the injury potential, V_0, in a single nerve fiber.
Shaded portion of the fiber represents injured region. R_1 represents
the inside resistance and R_0 the outside resistance.

be destroyed without affecting the other end, at least for many hours
and even days.   This destruction breaks down the membrane re-
sistance at one end and allows a current to flow freely between the
cytoplasm and the extracellular fluid.   Contact can thus be made to
the inside of the cell by placing an electrode on the injured portion of
the cell.   If now an electrode is placed on the uninjured portion of
the nerve fiber, a potential will exist between the two electrodes such
that the one on the uninjured portion is positive.   This is known as
the injury potential.   Whereas it is due to the membrane potential,
it certainly cannot be taken as a measure of it with any degree of ac-
curacy.

The relation between the injury potential and the membrane po-
tential is best understood in terms of the equivalent circuit, shown in
Figure 10.   A simple calculation will show that the injury potential,
V_0, is given by the relation:

$$V_0 = \frac{R_o}{R_o + R_i} E_m \tag{5}$$

where R_o is the resistance along the outside of the nerve between the two electrodes and R_i the internal resistance, which includes the resistance along the inside of the cell between the two electrodes as well as the membrane resistance. E_m is the true membrane potential.

It will be seen that $V_0 = E_m$ only in cases $R_o \gg R_i$. This condition is very difficult to fulfill. R_i is the internal resistance of the cytoplasm and cannot be altered. The outside resistance, R_o, is normally very low due to the large bulk of electrolytic solution outside the cell. This resistance can be increased some by decreasing the bulk of fluid, but it is difficult to make it even as high as R_i. For this reason, in nerve bundles it is practically impossible to measure an injury potential that is more than a fraction of the true membrane potential. In single nerve fibers it has been found possible to dry out the nerve enough so that R_o is apparently considerably higher than R_i. Under these very special conditions the injury potential probably approximated the resting potential (17).

In measuring these potentials it is especially important to make sure that the measuring device is not affecting the potential being measured. It will be seen from the equivalent circuit of Figure 10 that, if the effective input resistance of the measuring instrument is too low, R_o will effectively be made still smaller, thus increasing the inherent error of this method. It is necessary to check this very carefully, especially when working with single nerve fibers.

The action potential of nerve or muscle is the electrical indication of the passage of an impulse down the nerve or muscle. It can be measured by the same method as is used for measuring the membrane potential, *i.e.*, either directly by means of an internal electrode or by the injured-end technique. In addition, the "diphasic" action potential may be measured between two electrodes placed at different points along the uninjured nerve fiber.

A measurement of the magnitude of the action potential is subject to the same limitations and errors as have already been discussed for the resting potential. In addition, the potential changes take place very rapidly and it is necessary to make sure that the recording instrument responds rapidly enough. In careful work, it cannot be assumed even that the electronic amplifiers are responding rapidly enough, and the response time must be carefully checked in this case.

If one is interested only in the action potential, it is of course unnecessary to use a direct-coupled amplifier.

2. Electrical Potentials around Tissues

Electrical potential differences exist around single cells and tissues under special conditions, and they have been rather extensively studied by a number of investigators. No interpretation of these potentials has been agreed upon by the various investigators, but it seems clear that they must be interpreted on some basis different from that used in the case of the membrane potentials in nerve and muscle.

In general these potentials exist between different parts of a tissue, which is composed of a group of small cells. The potential must then normally exist between different sides of the single cells of the tissue. One can think of this situation in the case of skin, a gland cell, or a growing plant, where there is a definite polarity to the cells and where one would expect the two sides to be in contact with different solutions. It is well known that the membrane potential is a function of the concentration and composition of the fluid in contact with the cell membrane, so that, if two sides of a cell are in contact with different solutions, in general one would expect a potential to exist between the two sides. This is a possible explanation of these potentials in terms of quantities that are reasonably well understood. It is by no means the only possible explanation (see Lund, *18*).

Here again it would be necessary to draw an equivalent circuit of the particular system under investigation and assign at least approximate values to the circuit elements before an amplifier could be constructed intelligently for measuring these potentials. However, in general it could be said that, the higher the input resistance of the amplifier or measuring device, the less would be the possibility of introducing errors due to measurement.

E. IMPEDANCE MEASUREMENTS

1. Introduction

Not long after Ohm formulated the law that has come to be known by his name, it was recognized that the electrical resistance of any substance is one of its fundamental properties, and consequently this measurement was made on all manner of compounds and solu-

tions. When the measurement of the resistance of a biological ma-
terial was made, however, many difficulties were encountered that
had not been met in physical systems. The resistance of a tissue
was found to depend upon the magnitude of the current, the time it
had been flowing, and even upon the direction of flow. Further, if an
electromotive force were applied to a tissue and then removed, a po-
tential difference would persist for a long time thereafter. The situa-
tion was further complicated by the fact that all these phenomena de-
pended upon the composition and previous history of the electrodes
used. All this was completely contrary to Ohm's law and to the be-
havior of electricity in physical systems, so it was felt for a time that
a completely different set of laws would have to be used for biological
systems.

These difficulties did not prevent a number of investigators from
making measurements in an effort to obtain some empirical correla-
tion between resistance and structure or function. The effects of
such things as salts, injury, narcotics, physiological activity, etc.
were tried. In general it was found that the resistance of highly or-
ganized and oriented tissues such as skin or muscle is different in dif-
ferent directions, but that the resistance is decreased and is the same
in all directions when the tissue is dead.

In the latter part of the nineteenth century Kohlrausch intro-
duced the practice of using alternating currents for measuring the re-
sistance of electrolytic solutions in order to eliminate the electrode
difficulties. When this idea was tried on biological materials it was
found that the resistance was independent of the type of electrode
used and, for small currents, independent of the strength of the cur-
rent. Under these conditions, Ohm's law could be considered valid
and a logical interpretation could be attempted.

These measurements were made by means of a Wheatstone bridge,
and it was found that a true balance of the bridge could not be ob-
tained unless a capacitor was placed in an arm of the bridge adjacent
to the tissue. This meant that there was a capacitance associated
with the tissue, which disappeared with the death of the cells.

It was next found that the resistance of biological material, unlike
physical substances or electrolytic solutions, depends upon the fre-
quency of the alternating current which is used to measure it. Höber,
in 1912, showed that the resistance of blood is very much lower at
very high frequencies (*ca.* 10,000,000 cycles per second) than at low
frequencies. Further, he showed that, when the cells are killed, the

resistance is independent of the frequency and equal to the high frequency value. He interpreted these findings in terms of the Bernstein model of cell structure, and these arguments had a good deal to do with the acceptance of Bernstein's hypothesis as we know it today.

Bernstein postulated that cells consist of an ionized solution surrounded by a thin impervious membrane. Höber reasoned that at low frequencies the cells offer a very high resistance to the flow of current because of the high resistance membrane, and all current flow is through the electrolytic solution surrounding the cells. However, the cell membrane acts as a capacitance, which allows high frequency current to flow through it, so at high frequencies the current flows both through the inter- and intracellular fluids. This supposition also accounted for the capacitance known to be associated with cells. Since this work, a great many refinements of both measurement and interpretation have been added, but the fundamental concepts have not changed markedly.

2. Equivalent Circuit

The term impedance is used to designate the generalized form of resistance. The term resistance applies to direct currents and by Ohm's law is defined as the ratio between voltage and current. Likewise for alternating currents the impedance is defined as the ratio between the voltage and current, and may change as a function of the frequency. If the system contains no inductive or capacitative elements, the impedance is independent of the frequency and equal to the resistance. Thus the resistance is merely the zero frequency impedance of any system.

A living cell may be regarded as an impermeable membrane surrounding a conducting fluid, with the cell itself immersed in a conducting fluid. Thus two conductors are separated by an insulator, and this constitutes an electrical condenser. It has a certain capacitance, known as the membrane capacitance. However, the membrane is not a perfect insulator and some ions can pass through under the influence of an electric field. Thus there is a certain ionic conduction across the membrane. Since ionic conduction is independent of frequency, it is customary to speak of the ionic conduction phase of the membrane impedance as the membrane resistance.

Thus in general there are three paths that an alternating current can follow in flowing through a suspension of cells: (1) It can flow

through the intercellular fluid.   (2) It can flow through the intercel-
lular fluid to the cell, through the cell membrane by ionic conduction,
through the intracellular fluid, through the membrane on the other
side, and on to the other electrode.   (3) It can flow through the inter-
cellular fluid to the cell, across the cell membrane by induction by
virtue of the membrane capacitance, through the intracellular fluid,
out the other side by induction, and on to the other electrode.   On
this basis it is possible to draw the equivalent circuit, and this has
been done in Figure 11a.   It will be seen that the circuit of Figure 11a
can be reduced to that of Figure 11b by simple combination of series
and parallel elements.   This latter circuit is a reasonably close ap-
proximation to the impedance of all cells and tissues at all frequencies.
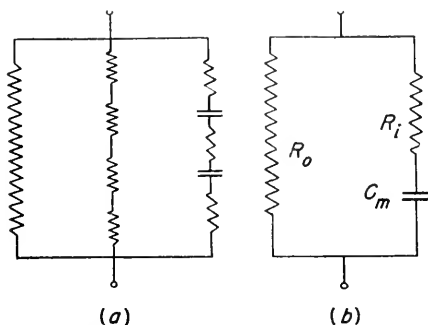
Fig. 11.   Equivalent circuit of a suspension of cells:
(a) complete circuit;   (b) reduced circuit.

The object of all impedance measurements then is first to assign exact
values to the three elements of Figure 11b, and then to interpret these
values in terms of the structure of the cell.   Neither objective is by
any means simple, and each will be considered separately.

3. Methods of Measurement

Complete discussion of the apparatus used for making impedance
measurements is beyond the scope of this book, and it will be possible
here only to indicate some of the major requirements that must be
met before impedance measurements are attempted.

First it is necessary to have a very accurate alternating current
Wheatstone bridge.   If only low frequencies of about a thousand cy-
cles are needed, it is possible to purchase a conductivity bridge, since
several good ones are available commercially.   Such measurements

may be quite useful for some limited studies. However, if at all possible the frequency should be variable over at least a limited range. Electrode troubles are best diagnosed by varying the frequency.

In order to obtain a complete and accurate measurement of the impedance of biological systems it is necessary to have a bridge with auxiliary oscillators, amplifiers, etc. capable of making accurate impedance measurements over a frequency range from a few cycles per second to several million cycles per second. Further, it must be capable of covering an impedance range from a few ohms to several hundred thousand ohms. No commercial equipment is available that will begin to do this, and the investigator must build his own. A complete description of a satisfactory bridge, oscillator, and amplifier has been published by Cole and Curtis (6). The impedance sensitivity of this bridge can be made 0.001% with an absolute accuracy over the entire frequency range of 0.1%. Such measurements should not be attempted with a bridge inferior to this, since often important quantities occur that are measured as small differences between much larger quantities.

In addition to the bridge it is necessary to have carefully designed conductivity cells for the particular application in question. At least one conductivity cell must usually be constructed for each problem undertaken. In general there are two problems involved in the design of a conductivity cell. First the conductivity cell must be of the proper size and shape so that current will flow through the tissue in a way such that the data obtained will be susceptible to analysis. Second, it must be designed in such a way that electrode effects are kept to a minimum.

Electrode effects are probably the most troublesome feature of these measurements and certainly at low frequencies place a definite limit on the accuracy attainable. There is always a certain resistance encountered in going from the electrode into an electrolytic solution and this resistance has a capacitance associated with it. These are known as electrode resistance and capacitance, respectively, and both are inversely proportional to the frequency. For this reason they are known also as polarization resistance and capacitance. At low frequencies it is easily possible to have the polarization capacitance far larger than the capacitance of the biological system. At higher frequencies the electrode effects become negligible. Polarization effects will be smaller the larger the resistance between the electrodes and the greater the area of the electrodes.

The type of electrode is enormously important here, and one electrode may exhibit polarization effects a thousand times greater than another.   The best electrodes are made of platinum and covered with a dense layer of platinum black.   This will minimize electrode effects but by no means eliminate them.   These electrodes are not reversible except under special conditions and therefore they may have a d.c. potential difference between them.   This causes no trouble for impedance measurements but means that in general it is not possible to use the same electrodes for potential and for impedance.   If it is necessary to use the same electrodes for both, silver–silver chloride electrodes are about as good a compromise as can be found.

4. Cell Constants as Measured by Impedance

The impedance of cells or tissues depends upon a good many different characteristics of the cells or tissue, and therefore we may use impedance measurements to determine some of these characteristics.

Volume Concentration of Cells.   This can be measured quite accurately by the impedance method provided the cells are of a known geometrical shape and are reasonably uniform.   It is obtained from the extrapolated zero frequency resistance, and the necessary equations have been worked out for the case of suspensions of spherical or spheroidal cells such as blood by Fricke (*14*), and for suspensions of cylindrical cells such as nerve and muscle by Cole and Curtis (*11*).   For these special cases, the accuracy for determination of volume concentration would probably be better than any other known method, up to volume concentrations so high as to cause deformity of the cells.   However, there are some inherent errors such as effects of stirring, surface conductance, etc. that limit the accuracy at the present time.   For tissues in general it can be said that the zero frequency resistance would be proportional to the volume concentration of cells but no quantitative relations have been worked out and indeed it is not to be expected that any great accuracy could be obtained by this method.

Thus by using a single low frequency, the impedance method can be used for the routine determination of volume concentration of blood or other cell suspensions, the fat content of milk, etc.   It has also been used for measuring the moisture content of soils, wood, etc.   A frequency of about 1000 cycles is high enough to minimize electrode errors in most cases and low enough to be practically equal to the zero frequency resistance of most cell suspensions or tissues,

Membrane Capacitance. This is a fundamental property of all cells and can be determined quite accurately by the impedance method. It has been worked out for a large number of different cells and surprisingly enough it amounts to about 1 μf. per square centimeter of cell membrane for all cells so far measured, and seems to be independent of the condition or state of functional activity of the cell (9).

Resistance of Interior of Cell. This can be measured quite accurately for cells of known geometric shape, and is computed from the extrapolated infinite frequency resistance. This is of course one of the important cellular constants and, in those cases in which it has been possible to check this method with other methods, the agreement has been very good.

Membrane Resistance. This is one of the most important properties of the cell, since it is presumably proportional to the ionic permeability; many attempts have been made, some successfully, to measure it by this method.   Unfortunately, this measurement is very difficult.   The reason for this can be seen from Figure 11a.   The membrane resistance is relatively high and is shunted by the very low resistance of the solution and by the low reactance of the membrane capacity.   In the case of a cell suspension the fact that there was an actual ionic flow of current through the membrane would become evident only as a discrepancy in the computed volume concentration.   In order to measure the membrane resistance then, the volume concentration would have to be measured by an independent means, and, if there is a difference between this and the volume concentration as computed from the impedance data, the membrane resistance could be computed.   However, since the best volume concentration measurements are good to only about 1% even in the most favorable circumstances, the membrane permeability would have to be relatively enormous before the impedance volume concentration would be appreciably smaller than that measured by other methods.   For this reason the membrane resistance has never been directly measured by this method on a suspension of cells.

In one instance, that of a single nerve fiber from the squid, a change of impedance was observed to accompany the propagation of the nerve impulse, and this was interpreted as being due to a change in ionic permeability of the cell membrane (8).   Even here no accurate absolute measure of the membrane resistance was obtained, but changes in resistance were recorded with considerable accuracy.

The membrane resistance has been measured with reasonable accuracy for several single cells. Cole and Hodgkin (7) measured the longitudinal impedance of the single nerve fiber from squid when immersed in mineral oil and Hodgkin and Rushton obtained reliable values in much the same way for a single nerve fiber from the crab. Blinks (5) has obtained values of the membrane resistance of single plant cells by impaling them on a microelectrode and measuring the resistance directly across the cell membrane. Thus impedance measurements can be used in very special circumstances for the estimation of ionic permeability.

Membrane Inductance. Under very special conditions it has been possible to measure an inductance associated with the cell membrane of the giant nerve fiber of squid. Whether inductance is associated with other cell membranes is not known. Even if it were, the problem of measuring it would present such difficulties that it is doubtful if it could ever be measured except in very special instances (10).

References

1. Abramson, H., *Electrokinetic Phenomena*. Chemical Catalog Co., New York, 1934.
2. Abramson, H., L. S. Moyer, and M. H. Gorin, *Electrophoresis of Proteins and the Chemistry of Cell Surfaces*. Reinhold, New York, 1942.
3. Bayliss, W. M., *Principles of General Physiology*. Longmans, Green, New York, 1927.
4. Bearden, J. A., *Rev. Sci. Instruments*, **4**, 271 (1933).
5. Blinks, L. R., *J. Gen. Physiol.*, **13**, 495 (1930).
6. Cole, K. S., and H. J. Curtis, *Rev. Sci. Instruments*, **8**, 333 (1937).
7. Cole, K. S., and L. A. Hodgkin, *J. Gen. Physiol.*, **22**, 671 (1939).
8. Cole, K. S., and H. J. Curtis, *J. Gen. Physiol.*, **22**, 649 (1939).
9. Cole, K. S., *Cold Spring Harbor Symposia Quant. Biol.*, **8**, 110 (1940).
10. Cole, K. S., *J. Gen. Physiol.*, **25**, 29 (1941).
11. Cole, K. S., and H. J. Curtis, *Cold Spring Harbor Symposia Quant. Biol.*, **4**, 73 (1936).
12. Curtis, H. J., and K. S. Cole, *J. Cellular Comp. Physiol.*, **19**, 135 (1942).
13. DuBridge, L. A., and H. Brown, *Rev. Sci. Instruments*, **4**, 532 (1933).
14. Fricke, H., *Cold Spring Harbor Symposia Quant. Biol.*, **1**, 117 (1933).
14a. Getman, F. H., *Outlines of Theoretical Chemistry*. 4th rev. ed., Wiley, New York, 1927.
15. Glasstone, S., *Textbook of Physical Chemistry*. Van Nostrand, New York, 1946.

16. Höber, R., *Physical Chemistry of Cells and Tissues*. Blakiston, Philadelphia, 1945.
17. Hodgkin, L. A., and A. F. Huxley, *J. Physiol.*, **104,** 176 (1945).
18. Lund, E. J., *Bioelectric Fields and Growth*. Univ. Texas Press, Austin, 1947.
19. MacInnes, D. A., *Principles of Electrochemistry*. Reinhold, New York, 1939.
20. Müller, R. H., R. L. Garman, and M. E. Droz, *Experimental Electronics*. Prentice-Hall, New York, 1945.
21. Penick, D. B., *Rev. Sci. Instruments*, **6,** 115 (1935).
22. Turner, L. A., *Rev. Sci. Instruments*, **4,** 665 (1933).

ELECTROPHORESIS

DAVID R. BRIGGS, *University of Minnesota*

A. CHARACTERISTICS OF THE METHOD. DEFINITIONS

The migration of particles, when suspended in a liquid, that is due to the influence of an imposed electric field is termed electrophoresis. The phenomenon is not essentially different from that of the migration of any ion in an electric field. It is dependent, fundamentally, upon the existence of an electric charge on the surface, or within the body, of the particle that is neutralized electrostatically by an equal charge of opposite sign, in the form of ions, distributed within the fluid adjacent to the particle, but at such a distance that, when placed in an externally imposed electric field, some degree of relative motion of these regions of opposite charges is possible against the viscous resistance to shear within the layer of the fluid that separates them in space.

The charge carried by a particle can arise from either of two sources. The material of the particle may contain chemical group-

ings that are ionizable when brought into contact with a liquid. For example, sodium proteinate in contact with water will ionize just as will any electrolyte in solution but the anions, in this case, are an integral part of the protein molecule (or particle) and constitute point charges on the surface of or within the body of the particle. The sodium ions, however, will be kinetically independent of the particle and, while held in the region of the particle surface by electrostatic forces of attraction exerted by the anions, will diffuse into the body of the liquid to a distance determined by the relative magnitudes of the forces of diffusion and electrostatic attraction. Other materials, examples of which are quartz or cellulose, which do not contain ionogenic groupings, will nevertheless acquire a charge when placed in contact with a liquid due to a differential adsorption by the solid surface of the electrolyte species present in the liquid. The hydroxyl ions present in water are more surface-active (more highly adsorbed) than the hydrogen ions. Surfaces of such ionogenically inert solids, then, usually bear a negative charge when in contact with water because of the statistically oriented state of the ions derived from water, wherein the OH^- ions are more strongly attracted toward the solid surface while the H^+ ions distribute themselves toward the water phase in the interface region.

As a result of this oriented distribution of the various ionic constituents at the particle surface, an electrical potential difference will exist between two layers of the fluid in the region of a particle–fluid interface. One fluid layer is considered to be fixed on the surface of the particle together with the fixed electrical charges of the surface, while the other, containing the counter electrical charges (counterions or gegenion), exists parallel to and a short distance normal to the particle surface and is considered to be free to move under any applied shearing force with respect to the fixed layer. This potential, which may exist in the region of any interface, is called the electrokinetic or zeta-potential. Its magnitude is dependent upon the net charge density per unit area of the layers, upon the distance apart in space at which the electrical centers of gravity of the layers exist, and, as a corollary to the latter, upon the dielectric constant of the medium occurring between the charged layers. The velocity of migration of a particle will be directly proportional to the magnitude of the electrokinetic potential at its interface and to the magnitude of the imposed field.

Helmholtz (1879–1888), in his early treatment of the theory of the

electrokinetic potential, considered both layers of the electrical double
layer to be planar layers of charges existing at a fixed distance apart
in the region of and parallel to the surface of the particle.   Guoy
(1910), and later Debye (1924), emphasized the necessity of consider-
ing the outer layer as an atmosphere of ions in which the density of
charge decreases from a large value near the surface to zero at a point
beyond the maximum distance that a counterion can reach kineti-
cally, while still being attracted electrostatically by the fixed charges
on the surface.   The effective thickness of the double layer is there-
fore to be considered as the distance of separation of the fixed charge
on the surface (including any charge elements within the layer of
fluid that is immovable with respect to the actual particle surface)
and the electrical center of gravity of the charged elements consti-
tuting the diffuse outer layer.   The distance from the particle sur-
face at which the change in potential with distance becomes zero will
be a function of the density of population of ions in the body of the
liquid, this distance decreasing in proportion to the ionic strength of
the solution bathing the particle surface.   The effective thickness of
the double layer, charge density on the particle remaining constant,
will decrease similarly with increase in the ionic strength.   On the
basis of this definition of the thickness of the double layer, the equa-
tions of Helmholtz may still be used in the interpretation of the
phenomena.   The thickness of the double layer and the magnitude
of the electrokinetic potential, however, cannot be regarded as fixed
and invariable at constant charge density but must be expected to
vary as a function of the ionic strength of the solution with which the
particle is in contact.

    The displacement per unit time interval that one layer of the
electrical double layer in the region of the interface will experience
with respect to the other when an externally applied electric field is
imposed on the system in a direction parallel to the planes of the
double layer will be a function of the forces, electrical and frictional,
acting upon the system.   When the volume element containing one
layer of the double layer is not capable of displacement in space,
while that containing the other layer is displaceable, all relative move-
ment observable will occur as a spacial displacement of the latter.
For example, if the double layer is one existing in the region of the
interface of a solid capillary filled with fluid and an external potential
is applied through the fluid in a direction parallel to the lumen of the
capillary, all relative movement of the two layers of the double layer

will occur with respect to the fixed wall of the capillary and will be observable as a flow of the liquid through the capillary. This phenomenon is called electrosmosis. If, however, the double layer is one existing in the region of the surface of a particle suspended in a fluid contained in a vessel in which no net displacement of the fluid can occur, all the relative movement observable between the two layers of the double layer will be imparted to the particle and it will migrate with respect to the liquid or any fixed point on the vessel containing the system. This phenomenon is called electrophoresis.

B. ELEMENTARY THEORY

The force that acts to cause displacement of a particle in electrophoresis is determined by the magnitude of the charge carried by the particle, and by the strength of the applied electrical field. If the field strength is taken in terms of the voltage drop per centimeter (E) and the net charge on the particle is taken in coulombs (Q), the force of acceleration (f) acting on the particle will be equal to the product of these quantities, i.e., $f = EQ$. As the particle is accelerated by the action of this force, it will meet a resistance to its motion through the liquid that is a function of the hydrodynamic flow characteristics of the region within the liquid in which viscous flow occurs as the liquid moves around the particle. Two limiting conditions may be treated in describing the magnitude of this force of viscous resistance that the moving particle will encounter within the liquid and with which the force of acceleration will quickly attain equilibrium to define the steady state velocity of the particle.

If the particle is described as a small sphere of radius r such that r is greater than the dimensions of the molecules of the medium of suspension (condition for validity of Stokes law) but small compared to the thickness, d, of the double layer ($r \ll d$), the particle may be treated as a charged sphere suspended in a uniform field and the force of viscous resistance, f', that it will encounter will be given by Stokes law, i.e., $f' = 6\pi\eta r u$. At the steady state, $f = f'$ and $EQ = 6\pi\eta r u$. Thus $u = EQ/6\pi\eta r$, where u is the velocity of the particle (cm./sec.), η is the coefficient of viscosity (poise) of the fluid through which the particle moves, and r is the radius of the particle. The mobility (m) of the particle (cm./sec./v./cm.) is equal to u/E or $m = Q/6\pi\eta r$. From this relationship it is apparent that for sufficiently small particles suspended in solutions of constant viscosity and of constant but very low ionic strength (where $d \gg r$) the mobility of the particle

is a function only of its net charge and of its radius, if spherical.
With small particles and low ionic strength of solutions it is to be
expected that shape of particle will also be a factor in determining
its mobility.

If the particle is of such dimensions that its radius of curvature is
large compared to the thickness of the double layer then, independent
of particle shape, the double layer may be considered as parallel
plates of zero curvature with respect to the applied field (*i.e.*, curva-
ture will be parallel to the contour of the applied field in the neigh-
borhood of the particle surface). The net charge carried by the par-
ticle, Q, can be resolved into the net charge density per unit of surface,
σ, by the relationship $\sigma = Q/A$, where A = the area of the particle
surface (cm.²). The force of acceleration (f) experienced per unit of
surface by the particle under field strength E will then be $f = E\sigma$.
All relative movement between the particle and the fluid must then
take place in the fluid existing between the planes of the double layer,
and a resistance to this movement will exist determined by the vis-
cosity of the liquid, the velocity of the particle, and the distance in
the liquid between the point of maximum motion (*i.e.*, that of the
layer fixed on the particle surface) and the point of no motion (*i.e.*,
beyond the outer layer of the double layer). If the viscous flow
of the liquid between these regions can be considered linear, as be-
tween two parallel plates moving with respect to each other, the
frictional resistance, per unit of particle–liquid interface, to relative
motion of the particle and the fluid phase will be given by the expres-
sion $f'' = \eta u/s$, where η is the viscosity coefficient (poise) of the liquid
in the region of the interface, u is the velocity of displacement (cm./
sec.) of the particle with respect to the liquid, and s is the distance
(cm.) between the regions of maximum motion and no motion. At
equilibrium, when the steady state is reached, wherein the force of
acceleration and force of resistance to motion are equal, $E\sigma = \eta u/s$
or $u = E\sigma s/\eta$ and $m = \sigma s/\eta$. Since the entire force of resistance is
exerted within the volume of liquid existing between the two layers
of opposite charge (*i.e.*, within the double layer) it can be assumed
that s will be equal to or proportional to the thickness of the double
layer and will vary with the ionic strength of the solution in the same
manner as will the thickness of the double layer. It is evident from
this relationship that the electrophoretic mobility of a particle sus-
pended in a solution of a given viscosity and ionic strength is a func-
tion only of its net charge density when the particle is large enough

and the ionic strength is high enough so that the thickness of the
double layer is small compared to the radius of curvature of the par-
ticle.   Under these conditions shape of the particle will not affect
its mobility.

It is understandable that, in order to calculate the value of the
net charge (or the net charge per unit of interface *i.e.*, charge density)
carried by a particle and the value of the electrokinetic potential
existing in the region of the interface from mobility measurements,
considerable knowledge must be available as to size and shape of the
particle and as to the theoretical relationship existing between the
thickness of the double layer and the ionic strength of the solution.
Such relationships are complex.   Fortunately, electrophoretic mo-
bility measurements can yield much valuable information without
the necessity of obtaining values of net charge or of the electrokinetic
potential therefrom.   In solutions of variable environment (variable
*p*H, ionic strength, etc.), for example, particles that migrate always
with equal mobilities may safely be considered to possess equal
charge density and zeta-potential characteristics and, therefore, can
to a high degree of probability be considered as identical in shape,
size, and surface characteristics, generally.   Particles that show dif-
ferent mobilities in contact with a solution of common environmental
characteristics must, in contrast, possess different surface properties
and cannot be of identical surface compositions   It is in this regard
that electrophoresis experiments have proved most useful in bio-
physics; this method constitutes one means of characterizing a given
biochemical preparation as to homogeneity or heterogeneity.   Thus,
from comparisons of the electrophoretic mobility alone, as obtained
under standardized experimental conditions, considerable informa-
tion can be obtained as to the physical nature of the charged particu-
late material being studied.   As a corollary, electrophoresis can serve
as a means of separating, in solution, materials of different surface
properties and can thus serve as a preparative method for substances
that might be difficult to separate from impurities by other means.

There are two methods in general use for the determination of the
electrophoretic mobility of charged particles dispersed in a liquid,
namely, the microelectrophoresis method and the macroelectrophore-
sis, U-tube, or moving-boundary method.   Each method has certain
advantages over the other and also certain limiting conditions under
which it may be most successfully employed.

As the name implies, the microelectrophoresis method employs a

microscope for following the movement of particles in the electric field. This method is primarily restricted to studies on particles that are large enough and possess a refractive index sufficiently different from that of the suspensions medium to render them visibly detectable, under light- or darkfield illumination, with a microscope.

In the moving-boundary method the displacement of the particles in an electric field is observed, in the legs of a U tube, as a change in position with time of a boundary existing between a solution containing the particles and another solution of, ideally, the same composition with respect to all components except the particulate component being studied. This method finds its greatest usefulness in the study of the mobilities of substances in solution of such particle size and refractive index value as to be undetectable under the microscope, though it may be employed on stable suspensions of larger particles if desired. The two methods may be regarded as supplementary with respect to the size range of particles that may best be studied by each but this is not necessarily the only or even the most important criterion upon which a choice of method will be based.

C. MICROELECTROPHORESIS METHOD

1. Applications and Relative Advantages

Measurement of the mobilities of particles such as blood cells, bacteria, oil droplets, and of finely particulate solids, such as quartz, glass, collodion, etc. can best be made with the microelectrophoresis method. The relative advantages of the technique, whenever it can be applied, may be listed as follows:

(1) It is the only way in which the electrokinetic properties of certain biological systems (e.g., microorganisms) can be accurately investigated.

(2) The environment in which the particle is observed does not change during the observation. This eliminates a condition of uncertainty associated with the moving-boundary method in which the values of pH, ionic strength, and individual ion concentration may vary considerably across the boundaries, i.e., the very point in the system where mobilities are being followed in the U tube.

(3) Mobility measurements may be made in solutions of very low ionic strength, a condition that vastly amplifies the uncertainties of the moving-boundary method. On the other hand, mobility measurements at ionic strengths above about 0.10 should not ordinarily be attempted with the microelectrophoresis method because of convection current disturbances in the

solution within the cell caused by the heat effects of the large electric currents that will pass at higher conductivities of the solution.

(4) The shape, size, and orientation of the particles may be directly observed and the mobilities of the various particles may be compared by watching them move simultaneously.

(5) Small quantities of disperse phase are required for a determination; in fact too high a density of particle population must generally be avoided in order to prevent confusion in the microscope field.

(6) It may be adapted to electrophoresis measurements in nonaqueous mediums.

(7) Measurements may be made through a wide range of pH, with a minimum of time consumed.

(8) A very distinct advantage lies in the fact that individual determinations require only a few minutes.

(9) The apparatus is inexpensive and relatively simple to set up and to use.

An important extension in the applicability of the microelectrophoresis method occurs in the case of soluble materials of such small dimensions as to be microscopically invisible themselves but which, being surface active, will accumulate on the surfaces of visible particles and impart to the surfaces of such particles the electrophoretic properties of the adsorbed substance. Particles of nonionogenic solids, such as finely ground quartz or finely dispersed collodion, which show a relatively high interfacial tension in contact with water, will adsorb many water-soluble, surface-active substances such as albumins, gums, soaps, etc. to such an extent that the electrical properties of the original solid-water interface are completely masked by those of the newly formed, adsorbed substance–water interface. The mobilities of such covered particles can then be determined microelectrophoretically and have repeatedly been found to agree closely with the mobilities of the same materials when in solution as measured by the moving-boundary method. This is an experimental fact that is not easily explained on the basis of electrokinetic theory, but it is a fortunate situation insofar as the experimenter is concerned since it makes possible the determination of the mobility properties of many substances by use of the relatively simple micro technique which otherwise could only be studied by the moving-boundary method, notably a more time-consuming operation. Isoelectric points of soluble, purified protein preparations are usually determined by this procedure. The investigator should be warned, however, that where mixtures of materials of varying degrees of surface activity are being

studied, the mobility properties observed may apply to only one (the most surface-active) component present or it may apply to no one component but to the equilibrium mixture of components (if all are of comparable adsorbability).   The relative effectiveness of the various components may well vary also with such conditions as pH and ionic strength or with the nature of the solid particle employed.   Such possible limitations of the method must always be considered in drawing conclusions from observations made on native mixtures.

2. Method of Observation

For the above reasons the microelectrophoresis method will usually be preferred for the study of the electrophoretic mobilities of solid or liquid particles of microscopic size, of monocellular organisms, or body cells.   It will find frequent application in studies on purified preparations of proteins or other soluble high molecular substances that are strongly adsorbable on microscopically visible particles.   A variety of apparatus has been devised for this purpose.   The essential features, which are common to all, of these apparatus, are: (a) an observation cell in which the solution containing the particles is placed and which is of such construction that a microscope can be focused upon the particles contained therein; (b) the ends of this observation cell are connected to electrodes through which an electric field can be introduced to the solution in the cell; and (c) some provision is usually made by which the cell can be filled and emptied without disturbing the assembly after it is mounted on the microscope stage.   These apparatus vary primarily as to shape (flat or cylindrical) and size of lumen of the cell, in the nature of the electrode systems employed, in the manner of assembly, and in the manipulation required during measurements.

In choosing a cell for microelectrophoresis work, a number of important characteristics that will determine the ease and accuracy with which measurements may be made need to be considered.   The mobility (cm.^{2}/v. sec.) of a particle is obtained from the simultaneous measurements of two quantities: (a) the velocity (cm./sec.) of the particle and (b) the field strength (v./cm.) acting on the particle. The velocity of the particle is obtained by observing with the aid of a calibrated eyepiece micrometer the displacement experienced by the particle in unit time as caused by the applied electric field.   It is essential, however, that this measurement of particle velocity be made

at a depth within the cell where the motion of the particle *with respect to the liquid* is the only motion observed.   Usually, in order to eliminate movement of the fluid in the cell from chance vibrational displacements, one end of the cell is tightly closed at the time of measurement.   Whether this is the case or not, when an electric field is applied along the lumen of the cell there will occur an electrosmotic displacement of the fluid along the interface between the cell wall and the fluid and, in order to maintain hydrostatic equilibrium in the vessel, a counterflow of fluid must take place in the center of the cell lumen.   Particles present will be carried along with these layers of moving fluid as well as be themselves displaced by electrophoresis in the electric field.   The result is that at only two regions within the cell, and along the line of focus of the microscope, will the observed movement of particles be due only to electrophoretic displacement. In a flat cell these regions of static fluid will be found at 0.211 and 0.789 of the depths of the cell.   In a cylindrical cell the static layer occurs at 0.147 of the diameter from the wall of the cylinder.   Measurements of particle velocity must be made at these depths in the cell. Because of the lenslike action that occurs at the wall of a cylindrical cell it is difficult to know accurately at what depth in the cell the microscope may be focused (but see alternative method on page 282). This is not so with the flat cell.   For this reason, the flat cell generally is preferred to other cell shapes.   The depths given as the ones at which measurements should be made in this shape of cell, however, are correct only if the width/depth ratio in the cell is greater than 20. Even then measurements should be made near the center of the width axis and not near the edges of the cell.   The movement of the fluid in the flat cell due to electrosmosis will follow a parabolic curve where displacement is a maximum in one direction at the walls and in the other direction at the center of the cell.   If the parabolic displacement curve is to be symmetrical in the cell and the 0.21 and 0.79 levels are to be correct, the electrosmotic displacement of fluid at each wall (top and bottom) of the cell must be equal.   Hence it is necessary, for the sake of simplicity in making measurements, that the top and bottom sheets of glass be of identical electrokinetic properties, *i.e.*, be constructed of glass of identical composition.

The thinner the cell the greater will be the rate of change in velocity of liquid flow due to electrosmosis with depth in the cell and the thinner will be the lamellae of liquid in which a minimum of motion of the fluid in either direction occurs.   It is necessary to be able to focus

sharply at the correct depth and to realize that the greater the diameter of the particle being observed the greater the probability that it will overlap above or below this exact depth and thus be given a plus or minus viscous drag by the moving lamellae of liquid above or below that of zero liquid flow.   The larger the diameter of the particles being observed the thicker should be the lumen of the cell for greatest accuracy.

The proper cross-sectional depth of the flat cell lumen and the optimal thickness of the glass plates making up the top and bottom of the cell will be determined roughly by the size range of the particles to be studied.   It is always necessary to be able to focus the microscope both on the top and on the bottom of the cell lumen in order that the 0.21 or 0.79 depths can be found.   The maximum optical distance from the bottom of the cell lumen to the top of the top plate must be less than the focal length of the microscope objective that may be required to resolve the particles.   Where very small particles are being followed and water or oil immersion objectives are needed, the lumen depth and cell wall thickness must be held at a minimum. Cells made from polished glass slides 0.5 to 0.6 mm. in thickness and assembled with a cell lumen depth of the same order of magnitude will be found usable with the smallest visible particles and will approach the limits for construction of such cells.   The cell should be of one piece of glass in order that the cleaning (with the alkali and chromic acid solutions) needed to render the cell walls free of contaminants may be easily carried out.   The all-glass cell also eliminates troublesome movements of liquid in the cell due to the vibrational or otherwise chance disturbances resulting from elastic distortions of rubber or other flexible tube connections.

The electrodes through which the electric field is applied across the ends of the cell must be nonpolarizing and of sufficient current-carrying capacity so that no gassing will occur.   The cell must be so constructed that no contaminating electrolyte from the electrodes will reach the point in the cell where measurements are being made, since even infinitesimal amounts of such electrolytes by changing the ionic strength of the solution will have profound effects in reducing the mobilities of particles, particularly when measurements are being made at low ionic strengths.   The cell should be so constructed that quick and easy change of the contents (of the flat portion of the cell where actual measurement is to be made) can be accomplished without danger of washing in electrolytes from the electrode regions.

3. Choice of Apparatus

Microelectrophoresis apparatus with observation cells of the cylindrical type (8) have the advantage of being simplest to construct. A simple capillary tube of 0.5 to 1.0 mm. bore, with an area ground and polished flat at the point where observation with the microscope is to be made, can be sealed to electrode chambers and to inlet and outlet tubes to complete the construction of the apparatus.   The difficulty of focusing at the proper depth for observation, however, seriously limits the accuracy of results obtained with the cylindrical cell.   A cylindrical cell (9) that employs two capillary tubes bearing definite ratios to each other both as to radius and length offers some improvement with regard to this criticism of cylindrical cells generally.   In this instrument back-flow of fluid resulting from electrosmotic displacement along the walls of the tubes is taken care of to
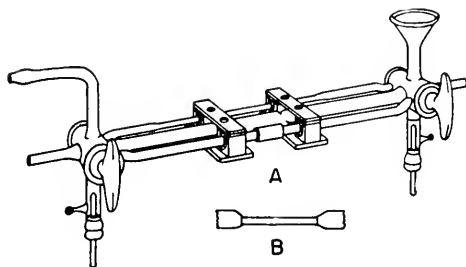
Fig. 1.   Flat horizontal microelectrophoresis cell of
Abramson and Moyer (2).

such an extent by the larger capillary (not the one in which observations are made) that there exists a static condition of the fluid in the center of the observation tube.   Under these conditions velocity measurements of particles made in the center of the small tube will be due to electrophoretic movement alone.   The greater increment of cell depth at which focus will be approximately correct allows for a higher probability of accuracy of measurement with this cell than is the case with a simple cylindrical cell.

While these and a number of other cells (10–13) have been used more or less successfully over the period of the last few decades, the flat cell instruments illustrated in Figures 1 and 2 have, in the experience of the writer and of many others who have used them, proved most amenable to routine laboratory use and approach most closely

the several requirements outlined above. That shown in Figure 1 is
the Abramson cell* (5,6) and is constructed of Pyrex glass, the flat
portion being made of polished Pyrex slides 0.6 mm. in thickness and
spaced 0.6 mm. apart. The cell† in Figure 2 is one devised by Briggs
(7) and is also constructed of Pyrex, the flat portion (A) being made
of polished Pyrex slides 1.0 mm. thick and spaced 0.8 mm. apart.

The Abramson cell employs Cu CuSO₄ electrodes. These are
connected to the flat portion of the cell through the lumens of right
angle bore stopcocks, which can be turned to other positions to allow
filling of the cell with the suspension to be studied. Care must be
taken to wash fluid through the stopcock lumens in such a direction
as to carry away any electrolyte that may have diffused into them
during preceding measurements, before fresh suspension is allowed
to pass into the flat cell.
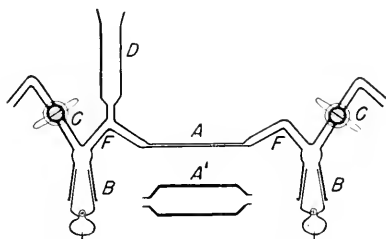
Fig. 2. Flat horizontal microelectrophoresis
cell of Briggs.

The Briggs cell uses Hg–HgNO₃–KNO₃ electrodes (B, Fig. 2),
which can be removed easily and cleaned and are connected to the
cell proper by standard taper, ground glass joints. This cell is so
constructed that fresh fluid containing the particles can be introduced
at any time (through D) without danger of contamination by elec-
trode electrolytes since the movement of this fluid will always be past
the electrode connection in a direction away from the flat part of the
cell. Flow of the test fluid past each electrode is governed by opening
or closing stopcocks (C) at the ends of the cell, and the outlets are so
arranged that, once the cell is filled, opening the outlet stopcocks can-
not result in complete evacuation of liquid from the cell proper.

* Obtainable from Hopf Glass Apparatus Co., 192 Third Ave., New York
City.
† This cell may be obtained from the University Glass Blowing Shop, Uni-
versity of Minnesota, Minneapolis, Minnesota.

The supporting frame used for mounting the Abramson cell on the microscopic stage is shown along with the cell itself in Figure 1. A simple frame for supporting the second cell can be made from two $1'' \times 7'' \times \frac{1}{8}''$ brass strips arranged parallel to the length of the cell and spaced and held $1\frac{1}{2}''$ apart by Bakelite blocks ($1'' \times \frac{1}{2}'' \times 3\frac{1}{2}''$). The latter are attached to the brass strips near their ends with screws and in such a position that these blocks can be shaped to fit, as saddles, the bends in the glass tubing of the cell (F, Fig. 2) on either side of the flat portion. These saddles are shaped to such a height that the flat cell portion is held in a horizontal position at the level of the bottom of the brass strips and the cell is anchored with thin metal straps to the support at the points of contact with the saddles. The whole assembly will set flat on the microscope stage with the flat part of the cell flush with the stage. Pins, put through one of the brass strips in such positions that they will fit into the holes in the microscope stage (usually occupied by slide clamps), will aid in keeping the cell and mount always in the same position on the stage and prevent accidental shifting from position after the proper depth in the cell has been attained for correct electrophoresis measurements. (*Note*: $1'' = 1$ inch $= 2.54$ cm.)

The Abramson cell as described has the advantage that somewhat higher degrees of magnification can be employed than is possible with the other cell because of the greater optical distance from the top of the upper glass plate to the bottom of the cell lumen in the latter. Each cell has some advantages over the other in ease of manipulation. Each will, in most circumstances, prove itself to be free of many technical disadvantages associated with most other cells described in the literature.

4. Measurements and Calculations

The mechanics of measurement of the electrophoretic mobility of particles by this method involves the observation of the motion of the particle at the proper depth in the cell, the measurement of the time required (using a stopwatch) for the particle to move a chosen distance, as indicated on a calibrated eyepiece micrometer, and a determination of the field strength acting upon the particle. The latter is best obtained from a measurement of the specific conductivity (λ) of the solution containing the particles, a measurement of the cross-sectional area (A) of the flat cell at the point of observation and a measurement of the current (I) passing through the cell at the time of measurement. The current is usually measured with a micro- or milliammeter placed in series with the electrophoresis cell in the circuit

supplying its electromotive force. Batteries of the "B" type supplying 45–90 v. can be used as the source of applied voltage. The field strength, E (v./cm.), acting on the particle will be given by the relationship, $E = I/A\lambda$, where I is in amperes, A is in square centimeters, and λ is in mhos. The mobility of the particle will then be $m = (d/t)/E$, where d is the distance (cm.) the particle travels in time t (sec.). m will have the dimensions square centimeters per volt second.

The value of λ, the specific conductivity, must be obtained for each suspension to be studied. This may be measured with a standard 1000 cycle Wheatstone bridge conductivity apparatus. The area of cross section of the cell, A, must be determined carefully for each new cell at the point along the length of the cell at which subsequent observations of mobility will be made. The average cell lumen depth is obtained by measuring the distance between top and bottom of the lumen at several points along the wide axis of the cell (at the correct position along its length) with the microscope micrometer focusing adjustment. The width of this axis is obtained by measuring with a microscope cathetometer.

When a new cell is put into use, it is desirable to test it for uniformity of construction by introducing into the cell a test suspension of quartz or collodion particles in a dilute solution (about 0.01%) of some surface-active substance such as a soluble protein (at some pH away from its isoelectric point) and measuring the velocity of displacement of the particles (under constant field strength) at a number of depths in the cell. When the observed particle velocity is plotted against the fractional depth in the cell at which the observation is made, a parabolic curve, symmetrical about the center, should be obtained. Because the walls of the cell will be coated with the protein just as the particles are coated, the electrosmotic displacement of fluid along the wall should equal the electrophoretic displacement of the particles there (but in opposite directions) and it should be found that u, the observed velocity of the particles, at the wall will be zero. Also under these conditions u at the 0.21 and 0.79 levels in the cell should be equal to $\frac{2}{3}$ u observed at the center of the cell (at the 0.5 level). If these relationships are found, the construction of the cell (width to depth ratio, parallelism of the top and bottom plates, etc.) is correct and the cell should yield dependable mobility measurements. As a further check on the dependability of the mobility measurement (which will be dependent upon the accuracy with which λ, A, I and u

have been made), which the experimenter is able to obtain, a freshly made dilute suspension of human erythrocytes in 0.067 M phosphate buffer of pH 7.4 can be observed in the cell. At 25°C. the observed mobility for these cells should agree very closely with the value 1.3 $\times$ 10^{-4} cm.² per volt second.

Since the microelectrophoresis cells are not readily constructed in a form susceptible to thermostatic control, it is always necessary to keep the wattage dissipation within the flat part of the cell at such a low value that convection currents are not generated. This is not difficult to attain when the electrical conductivity of the solution is very low but, as the ionic strength approaches 0.1, disturbances from this source will be encountered and measurement may become impossible. Heating of the contents of the cell from the microscope lamp must be avoided for the same reasons. Such heating can usually be avoided by placing a cell filled with water in front of the lamp so that heat rays will be absorbed before reaching the microscope system.

While a mixture of visible particles of varying mobilities can be recognized and, perhaps, statistically analyzed as to mobility distribution, this method cannot give any information as to the number or relative amounts of components in a nonmicroscopically visible mixture of disperse components (such as blood serum). It is in this regard, particularly, that the microscopic method loses preference to the moving-boundary method.

5. Some Recent Accomplishments

The microelectrophoresis method of measuring the mobility properties of microscopically visible particles has been applied to a wide variety of substances ranging from inorganic suspensions and emulsions to living cellular materials. Only a few examples need be referred to in order to illustrate the diversity of systems upon which such studies can yield important information. Red blood cells of humans are found to be remarkably constant in mobility (independent of sex or race) and to present surfaces not affected by other soluble proteins in the blood (14). Soaps are shown to impart mobilities (and surface charges) to dirt particles that are proportional to the detergent power of the soap (15). Latex particles are shown to be covered by proteins in some species and by lipides in others (16). Isoelectric points of soluble proteins and their change in charge with pH can be studied by this method (17). The nature of the surface membranes of micro-

organisms (18) and the effects thereon of various surface-active agents
(19) can be elucidated by such methods.   The interaction of proteins
with high affinity anions such as that of metaphosphoric acid can be
studied and the nature and extent of the interaction with changing
pH and anion concentration can be followed by microelectrophoresis
(20).

D. MOVING-BOUNDARY METHOD

1. Applications and Relative Advantages

The mobilities of substances that are submicroscopic in particle
dimensions and are not dependably adsorbable on larger microscopi-
cally visible particles may be studied only by use of the moving-bound-
ary technique.   Examples are the many lyophobic colloidal sols such
as gold sol, arsenic trisulfide sol, etc.   The moving-boundary method
is unique also in that it can yield information on the mobility charac-
teristics of mixtures of substances such as native solutions of proteins
(e.g., blood serum).   Some of the advantages and realms of best ap-
plication of this method may be summarized as follows:

(1) Of greatest importance is the capacity of the moving-boundary
method, used in conjunction with the refractive index methods employed for
the detection of the boundaries, to yield information as to (a) the electro-
phoretic homogeneity or heterogeneity of the disperse phase in an unknown
solution, (b) the number of components in a heterogeneous disperse system
that are electrically separable, (c) the degree of homogeneity of each compo-
nent, (d) the mobility of each component, and (e) the relative concentration of
each component in such a mixture.

(2) This method is applicable to a wide variety of high or low molecular
substances that form solutions in which there are no microscopically visible
particles.

(3) Electrophoresis measurements can be made in solutions of consider-
ably higher ionic strength than is generally feasible with the microscope
method.   The upper limit of ionic strength that can be employed in either
method is governed by the wattage dissipation in the region in the cell where
measurements of mobility are being made and the efficiency with which the
heat so generated is removed, i.e., the efficiency with which convection cur-
rents are avoided in that region of the cell.   In the U-tube method the cell
can readily be thermostated.

(4) Mobilities of a given substance as a function of pH may be studied
throughout the whole aqueous pH range.

(5) Interactions of one disperse component with another in solution may be detected and studied by this method.

(6) An advantage this method has to offer the biochemist is the possibility of separation in the pure state (or a purer state) of electrophoretically different components of a mixture.

Some disadvantages of the moving-boundary method as compared to the microscopic method are:

(1) The method requires a much longer time for individual measurements.

(2) Larger quantities of disperse substances are usually required. Generally the concentration of disperse phase should be 0.5 to 2.0% and ordinarily about 25 ml. of such a solution is needed to wash out and fill the U tube. This is distinctly a limiting factor in many cases of substances obtainable only with great difficulty. Newer cells and methods will tend to reduce this requirement. (Cells requiring not more than 20 mg. of disperse phase are already in use.)

(3) Conductivity and pH differences that exist across the boundary between the buffer and the solution containing the disperse phase to be studied lead to boundary uncertainties and anomalies that are not always resolvable. This is particularly true when the buffer solution employed is of low ionic strength (0.05 or less), and may lead to considerable difficulties in the interpretation of observations both as to the true mobilities of the components and as to their relative concentrations in the mixture. This constitutes the primary limitation of this method at the present time.

(4) The moving boundary method apparatus is more expensive than the microelectrophoresis apparatus.

2. Essential Requirements

Many variations of apparatus of the U-tube type have been devised from time to time (21–23) for the purpose of measuring the electrophoretic mobilities of a variety of colloidally dispersed substances by the moving-boundary method. Since the measurement of the mobility of a given component by this method is obtained by observation of the displacement, under the influence of an electric field, of a boundary between two solutions within the fluid system, only one of which contains the component in question, it is necessary that the following requirements be fulfilled in order that quantitative measurements can be accomplished: (a) The composition, pH, conductivity, density, etc. of the fluids in the region of the boundary must be unaffected during the course of the experiment by any elec-

trolysis products that may form at the electrodes through which the electric field is introduced into the system containing the boundary. Also any disturbances that could be transmitted to the boundary due to gas formation at the electrodes must be avoided.   (b) The apparatus must be so designed that a sharp boundary can be formed initially between the solution containing the component to be followed and a solution that does not contain this component.   (c) Disturbances of the boundary due to convection currents in the boundary region arising from wattage dissipation in the fluid column containing the boundary must be obviated.   (d) It must be possible to detect accurately the position of the boundary, visually or otherwise, both initially and after a measured lapse of time during which a constant electric field is maintained throughout the column of fluid containing the boundary.   (e) A means must be available for measuring and maintaining constant the imposed electrical field strength under which migration of the boundary occurs.   (f) The electrolyte composition, pH, and specific conductivity of the two solutions that meet at the boundary must always be as nearly identical as possible while still maintaining the disappearance in the boundary of the component to be studied.   As shall be shown, this requirement is the most difficult to approximate and is the source of the major limitation to the use of this method.

3. Modern Apparatus and Technique

The series of moving-boundary apparatus that have been devised actually constitute progressive stages of development in which the objective has been to meet more and more effectively the various requirements for quantitative measurements by the method.   An apparatus devised by Tiselius (24,25) represents such a great improvement over earlier instruments that these are now primarily of historic interest only.   Tiselius introduced the use of the refractive index or *schlieren* method for the detection of the position of the boundary in the U tube and also emphasized the importance of avoiding convection currents in the regions of the boundaries; he recommended methods by which these currents can be minimized.   The use of the refractive index method for characterizing the boundary has made possible also the analysis of the contours of the concentration gradients occurring in the boundary region and the analysis, therefore, of the electrophoretic homogeneity or heterogeneity of the material that disappears in the boundary.   Tiselius incorporated in his apparatus

many of the improvements that had appeared in earlier apparatus. His original apparatus has been further improved by Longsworth and MacInnes (26).   Modifications in the manner of detecting and recording the refractive index changes that occur through the boundary have been made by Longsworth (27), Svensson (29), Philpot (28), and others.   A description of the Longsworth-MacInnes modification of the Tiselius apparatus will serve to illustrate how and to what extent the requirements enumerated above have been met in modern moving-boundary electrophoresis apparatus.   A detailed description of the installation and use of this instrument has been given by Longsworth (30).
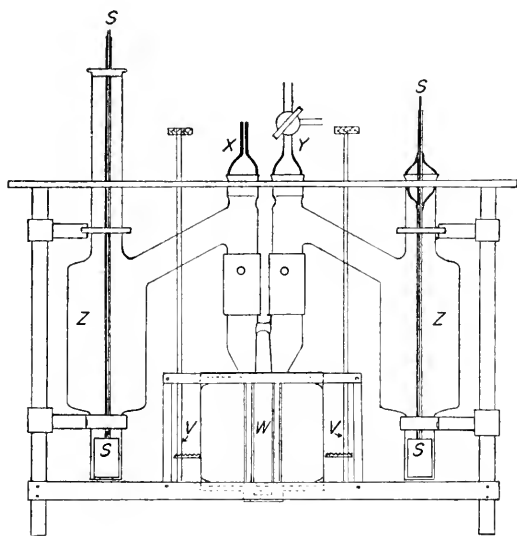
Fig. 3.   Moving-boundary electrophoresis cell of Tiselius with supporting frame, electrodes, and electrode vessels, showing mechanical devices for moving segments of the cell as required during filling of cell and formation of boundaries.
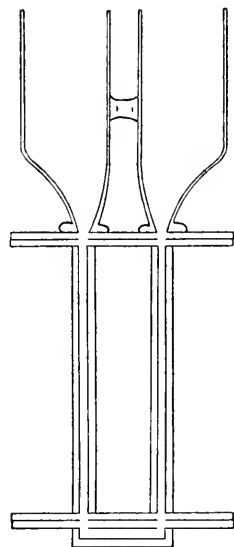
Fig. 4.   Analytical cell showing sectional construction that makes possible formation of sharp boundaries.

In Figure 3 is shown, diagrammatically, the construction of a modified Tiselius U-tube apparatus (with supporting frame) and Figure 4 is a drawing of the cell proper, i.e., the U tube in which the boundaries are formed and allowed to migrate electrophoretically.   As illustrated in Figure 3, this cell (W) is connected through rubber

sleeves (O) with large electrode chambers (Z) that contain silver–
silver chloride electrodes (S) which, when surrounded by a potassium
chloride solution, will allow the electric current to enter the system
without any accompanying polarization of the electrodes and attend-
ant evolution of gas. These electrode vessels are filled with a buffer
of the same composition as the overlying liquid at the boundaries.
This arrangement, and the long path existing between the electrodes
and the boundaries in the U tube, will give adequate protection against
the possibility that the products of electrolysis that accumulate in the
regions of the electrode will ever reach the region of the boundaries.
Requirement a is thus adequately fulfilled in this apparatus.

The cell proper shown in Figure 4 is one employed for analytical
purposes and consists of three sections. These sections are contacted
through ground glass plates lubricated by a grease that prevents leak-
age out of or into the cell (when the instrument is placed in a water
bath) and will allow horizontal displacement of parts of the cell with
respect to each other. This is necessary in order to obtain a sharp
initial boundary between the buffer and the solution containing the
material to be analyzed. Mechanical devices (V) are attached to the
frame and accomplish this required displacement of the center and
bottom portions of the cell when the cell is connected to the electrode
vessels through rubber sleeves attached to the static top portion of the
cell.

Filling of the cell may be accomplished as follows: With the cell and
electrode vessels assembled in the frame and the glass parts X and Y (Fig.
3) removed, protein solution (for example) is introduced into the U tube in
an amount sufficient to a little more than fill the bottom part of the U tube.
This bottom section is moved to the left, let us say, sufficiently to close the
lumen of the U tube. Then the right hand leg of the U tube is filled with
protein solution. Next, the left hand leg is washed with buffer and filled
with buffer. The center part of the U tube is now moved to the right closing
the connection of the lumen with the top portion of the cell. Both sides of
the top portion are washed with buffer, then filled with buffer together with
the electrode vessels. The glass parts X and Y are replaced, the electrodes
introduced, and saturated potassium chloride solution is passed into the
bottom of the electrode vessels, through the hollow tubes leading down to
the electrodes, until the electrodes are surrounded by the potassium chloride
solution. The electrode vessels are filled to an even level (top of piece X)
with buffer solution and the whole assembly is placed in a constant tempera-
ture bath. After the contents have reached bath temperature and with the
stopcock on part Y closed, the parts of the U tube are brought back into line.

This results in the formation of sharp boundaries between the protein solution and the overlying buffer at the top of the right hand leg and at the bottom of the left hand leg of the U tube. Gravitational equilibrium is attained between the solutions because the heavier protein solution occupies the bottom portion of the tube. The boundaries are displaced into view by slowly pumping buffer into the right hand half of the apparatus through the stopcock of part Y. After bringing the boundaries by this method into a convenient position in each leg, the right hand side of the apparatus is again closed, the current source is connected to the electrodes (making the left hand electrode that of opposite sign of charge to the sign of charge carried by the protein), and electrophoresis is allowed to proceed.

Thus requirement b is neatly accomplished with this apparatus.

In order to reduce convection currents due to wattage dissipation to a minimum, Tiselius recommended that the legs of the U tube be made thin in one dimension and that the temperature at which the electrophoresis is carried out be below the temperature of the maximum density of the solutions involved. For pure water this temperature is about $4°C$. and for most salt solutions (buffers) it will be above $0.5°C$. up to concentrations of 0.2 M. A convenient temperature to keep the water bath is $0.5°C$. With this arrangement and for a U tube of approximately 2.5 mm. diameter in the thin dimension, a wattage dissipation of about 0.2 watt per centimeter of tube length per square centimeter of tube cross section has been found to cause no convection disturbance when the density difference at the boundary is that due to a change in protein concentration of 0.5%. Requirement c is thus adequately fulfilled by this apparatus for aqueous systems of ionic strengths below approximately 0.2.

The method introduced by Tiselius for observing and characterizing the boundaries is the most novel feature of the apparatus. While a boundary may be detected visually in the case of colored sols, such as gold sol or As_2S_3 sol, there are many colorless colloid systems. By using U tubes made of quartz, under ultraviolet light, the boundaries in such systems as those containing proteins may be rendered visible by fluorescence or detectable photographically through absorption by the protein of the ultraviolet light. Often the boundary can be made visible by differences in the Tyndall effect. None of these methods has proved as versatile, however, as the method based on the refractive index changes that occur at the boundary between an overlying buffer solution and a protein solution, for example, equilibrated through a dialyzing membrane against the buffer. When light

passes through such a region of changing refractive index, its path
will be bent toward the region of higher refractive index. The index
of refraction of the protein solution will be greater than that of the
overlying solution by an amount proportional to the concentration of
the protein and to its refractive increment.

Figure 5 shows diagrammatically the manner in which this prin-
ciple may be applied to the detection of the position of a boundary in
the Tiselius cell. A horizontal slit, source of light (S) is allowed to
pass through a long focal length (schlieren) lens (L), which forms an
image of the slit at point P. If the light in passing from the lens to
point P encounters no region of refractive index change (existing in a
direction perpendicular to its direction of propagation) all the light
will be brought to focus as a simple image of the slit at P. If, how-
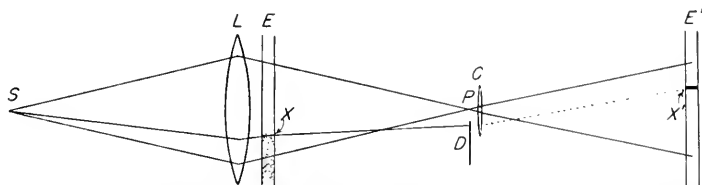
Fig. 5. Schlieren method for detecting position of a boundary (refractive in
dex gradient) in the macroelectrophoresis cell.

ever, an electrophoresis cell (E) is placed in the path of the beam from
L to P and this cell contains a boundary at X across which there is a
refractive index change in a vertical direction, that fraction of the
light passing through this region of refractive index change will be
bent down and will be brought to focus in the same plane as P but at
a point vertically below P. If a camera of long focal length is focused
on the electrophoresis cell with the camera lens (C) placed just beyond
the point P in the path of the light beam, all the light passing the cell
will be brought to focus on the camera plate. When a diaphragm
(D) is raised in the plane of P to a point where it will intercept the
light thrown down but not the main beam, there will appear on the
camera plate a shadow (X') in the image of the cell (E') corresponding
to that region containing the refractive index gradient, i.e., the region
of the boundary. Thus the position of the boundary in the cell can
be determined from the position of the shadow in its image. (See
also Figure 6 of Chapter III.)

The degree to which the light beam passing through the boundary

region is thrown down at point P will be proportional to the rate of change of the refractive index, n, with distance, x, within the cell. This rate of change of the refractive index, dn/dx, will be proportional to the rate of change of protein concentration, dc/dx, in the region of the boundary, varying as we pass from pure buffer through the boundary into uniform protein-containing solution, from zero up to a maximum and down to zero again. Thus, as the diaphragm (D) is raised, the width of the shadow (X') will increase. If we plot the rate of change of n with distance, dn/dx, in the cell, against the position, x, in the cell, a graph of the form shown in Figure 6 will be obtained for a symmetrical boundary. Such a figure is obtained automatically if
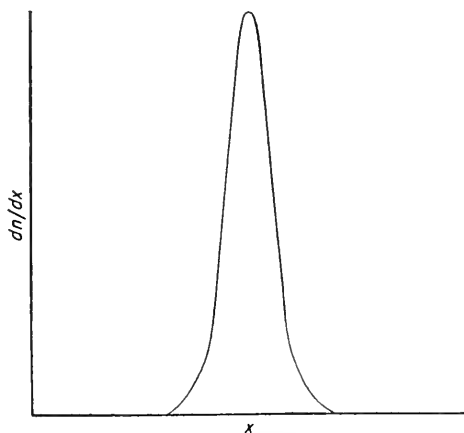
Fig. 6. Curve showing change of refractive index (n) with distance (x) in the cell, *i.e.*, dn/dx, plotted against position along the cell, x.

the photographic plate of the camera is moved uniformly across the cell image as the schlieren diaphragm is raised. This scanning method for obtaining dn/dx versus x patterns was introduced by Longsworth. Such curves describe the total refractive index change across the boundary and thus the total concentration change in protein across the boundary. The area under the curve is, therefore, proportional to the protein concentration change at the boundary. In the region of the boundary, dn/dx increases to a maximum (at the center of the boundary, if symmetrical) and decreases to zero in the region away from the boundary. If more than one electrophoretically different protein is present, a scanning picture taken after the

electric current has been allowed to pass for a time will indicate, as separate peaks in the pattern, the presence of more than one boundary. The number of boundaries observed corresponds to the number of electrophoretically different components present in the preparation. The area under the curve for each boundary compared to the sums of the areas for the total boundary will yield the relative amount of each such component in the mixture. Thus, this method can be used to analyze a protein preparation as to the number and relative amounts of electrophoretically different and independent components present and as to the electrical mobility of each. Requirement d, above, is very adequately met in this apparatus.

Requirement e in the above list, $i.e.$, the means for estimating the voltage drop per centimeter or field strength under which the boundary is migrating, can be approximated best by the method already referred to in the discussion of the microelectrophoresis method. If the legs of the U tube are of uniform cross section and the area, A (cm.²) of cross section is measured, the voltage drop per centimeter, E, in the cell will be given by the relationship $E = I/(\lambda A)$, where I is the current density (amp.) passing through the system (which can be calculated from the voltage drop occurring across a standard resistance placed in series with the electrical circuit to the cell) and λ is the specific conductivity (mho) of the cell contents in the region of the boundary. It is here, however, that an important uncertainty associated with the moving-boundary method arises.

4. Limitations

Because the two solutions that meet at the boundary cannot be of identical composition, it is to be expected that the two will vary to a more or less marked degree in conductivity, pH, and buffer salt concentration as well as in the concentration of the component that disappears in the boundary and whose electrophoretic properties are being investigated. These circumstances give rise to certain boundary anomalies such as (a) the nonmoving concentration gradients that appear at the site of the initial boundaries (the so-called δ and ϵ peaks in the electrophoresis patterns), (b) a difference in the distance traveled, in unit time by the rising boundary and by the descending boundary, and (c) differences in the shapes (sharpness) of the rising and descending peaks.

That the mobilities calculated from the distances moved by the rising boundary (where the colloid component is moving into the

region originally occupied by the overlying liquid) and by the descending boundary (where the colloid component is moving into a region originally occupied by this component and leaves behind a region, devoid of this component, originally occupied by the colloid solution) were often of quite different magnitudes has long been recognized. The question as to which, if either, constitutes the true mobility has been the subject of considerable discussion. It was not, however, until the refractive index method was used for characterizing fully the concentration gradients that develop in the boundary region that these gradients could be well enough characterized so that a possible quantitative treatment of the phenomenon could be given. At the descending boundary, where the colloid component is moving into a region of identical composition, these boundary uncertainties appear to be of less importance than at the rising boundary. It is a common procedure, therefore, to calculate mobilities from the boundary displacement observed in the descending boundary, using the value of λ obtained on the colloid-containing solution in the calculation of E.

The question as to what should be the proper or most suitable overlying solution for making contact with the colloid-containing solution at the boundary is still somewhat a matter of conjecture. Burton early suggested the use of an overlying solution adjusted to the same specific conductivity as that of the sol. This procedure eliminates, initially, the potential gradient that would otherwise occur at the boundary but it may prove only a momentary advantage due to diffusion and other adjustments that must subsequently take place. Kruyt and van der Willigen advised the use of an ultrafiltrate of the sol as the overlying liquid. This or its approximate equivalent, the equilibrium dialyzate of the sol against its buffer, is the solution that seems preferably employed in that it approximates most closely the desired minimum of environmental change for the sol component as it moves from the original boundary regions. In general, the higher the ionic strength of the buffer used, the lower will be the extent of occurrence of the anomalies rising from this boundary dilemma. Longsworth and MacInnes (*36,37*), Dole (*38*), and Svensson (*39*) have discussed the behavior of some typical systems with respect to these boundary uncertainties.

When mixtures of substances are being analyzed by the moving-boundary method it is sometimes found that the relative areas under the various peaks (denoting electrophoretically distinguishable components) may vary with conditions of the experiment such as pH, ionic

strength, type of buffer salts, and concentration of colloid components. Such variations will probably be due to interactions of the colloid components with each other or with salt components in the solution. In order for the analytical patterns to yield true values, *i.e.*, in order for the areas under the various peaks to truly denote relative concentrations of the various components, it may be necessary to search for conditions under which such interactions are minimized. While no generalizations concerning this phenomenon can be given as yet, it is mentioned here as a precaution against a too ready acceptance of any given pattern as representing a definite and invariable analytical relationship for the mixture involved.

In view of the importance of electrosmosis effects in the microelectrophoresis method, in which measurement of electrophoresis must be made at a definite depth in the observation cell in order to avoid errors due to electrosmotic displacement of the liquid along the wall of the cell, it might be expected that some precautions would be required in the moving-boundary method in order to take similar electrosmotic effects into account. While such displacements of the fluids along the walls of the U tube undoubtedly occur, experiments all seem to indicate that they cause no observable effects upon the positions of the boundaries. This probably results from the circumstance that the density differences that exist across the boundary serves to make this plane in the fluid system limiting with respect to electrosmotic movements that will occur within either solution meeting at the boundary. That is, circulation of the fluid in either solution arising from electrosmosis does not penetrate into the body of the other solution but is stopped there, the return flow in each being restricted to that solution. This circulation, with the boundary as the limiting plane, may promote a slight mixing there so that both rising and descending boundaries broaden out faster under the influence of the electric field than would result from the diffusion process alone. It is probable, also, that in any case in which the difference in density of the two solutions is small (*e.g.*, dilute colloid solutions) disturbances due to electrosmosis would cause inordinately high degrees of mixing of the solutions and could entirely invalidate the method. Experience has indicated that for this reason, as well as for the reason that the refractive index method for boundary detection requires appreciable concentration differences at the boundary, the moving boundary method should not be employed on solutions of less than about 0.25% of colloid component.

5. Some Recent Accomplishments

The perfecting of the U-tube method for electrophoresis measurement has enormously increased the usefulness of this electrokinetic technique in the study of materials of biological origin.   Since this method constitutes a means for analyzing proteins and other colloid electrolytes for electrophoretic homogeneity of components as well as for the number and relative amounts of such components present in mixtures, it can be used to characterize naturally occurring preparations, such as the proteins of the blood, in which variations from normal have proved of diagnostic value in some cases (*2*).   Proteins previously considered as single entities, such as casein, have been shown to be mixtures (*40*).   Even crystalline egg albumin has been shown to consist of two electrophoretically slightly different components (*41*).   The interaction of proteins and detergents can be studied by this method (*42*).   Some phases of the denaturation reactions of proteins are being elucidated by this means (*43*).   The usefulness of this method in helping to better characterize biocolloids becomes increasingly evident with each new system on which it is employed. The moving-boundary method has been adapted for use as a preparative procedure for the separation of electrophoretically distinct components from complex mixtures of natural occurrence (*31–35*). Successful purification of cytochrome c and the yellow enzyme are examples of the application of this method.

A complete analytical and preparative apparatus of the moving-boundary type employing the refractive index methods for boundary detection and analysis is obtainable from Klett Manufacturing Company, New York City. At the present time (1949), several more compact and simplified analytical instruments, designed primarily for clinical laboratories, are becoming available from various instrument manufacturers.

References

General References

1. Abramson, H. A., *Electrokinetic Phenomena*. Chemical Catalog Co., New York, 1934.

2. Abramson, H. A., L. S. Moyer, and M. H. Gorin, *Electrophoresis of Proteins*. Reinhold, New York, 1942.

3. Symposium, "The Electrical Double Layer," *Trans. Faraday Soc.*, **36**, 1 (1940).

4a. Symposium "Electrophoresis," *Ann. N. Y. Acad. Sci.*, **39**, 105 (1939).

4b. Moore, D. H., "Electrophoresis," in *Physical Methods of Organic Chemistry*, 2nd ed., A. Weissberger, ed.   Interscience, New York, 1949, Chap. XXVI.

Microelectrophoresis Method

APPARATUS AND TECHNIQUE

5. Abramson, H. A., *J. Gen. Physiol.*, **12**, 469 (1929).

6. Moyer, L. S., *J. Bact.*, **31**, 531 (1936).

7. Briggs, D. R., *Ind. Eng. Chem., Anal. Ed.*, **12**, 703 (1940).

8. Mattson, S. E., *Kolloidchem. Beihefte*, **14**, 309 (1922);   *J. Phys. Chem.*, **37**, 223 (1933).

9. Smith, M. E., and M. W. Lisse, *J. Phys. Chem.*, **40**, 399 (1936).

10. Northrup, J. H., and M. Kunitz, *J. Gen. Physiol.*, **7**, 729 (1925).

11. von Buzagh, A., *Kolloid-Z.*, **48**, 33 (1929).

12. Bull, H. B., *J. Phys. Chem.*, **39**, 577 (1935).

13. Kruyt, H. R., and A. E. van Arkel, *Kolloid-Z.*, **32**, 91 (1923).

APPLICATIONS

14. Abramson, H. A., *J. Gen. Physiol.*, **12**, 711 (1929).

15. Urban, W. M., and L. B. Jensen, *J. Phys. Chem.*, **40**, 821 (1936).

16. Moyer, L. S., *Am. J. Botany*, **21**, 293 (1934).

17. Abramson, H. A., *J. Gen. Physiol.*, **15**, 575 (1932).

18. Moyer, L. S., *J. Bact.*, **32**, 433 (1936).

19. Dyar, M. T., and E. J. Ordal, *J. Bact.*, **51**, 149 (1946).

20. Briggs, D. R., *J. Biol. Chem.*, **134**, 261 (1940).

Moving-Boundary Method

APPARATUS AND TECHNIQUES

21. Kruyt, H. R., and P. C. v. d. Willigen, *Kolloid-Z.*, **44**, 22 (1928).

22. Engel, I., and W. Pauli, *Z. physik. Chem.*, **126**, 247 (1927).

23. Svedberg, T., and A. Tiselius, *J. Am. Chem. Soc.*, **48**, 2272 (1926).

24. Tiselius, A., *Trans. Faraday Soc.*, **33**, 524 (1937);   *Kolloid-Z.*, **85**, 129 (1938).

25. Tiselius, A., and H. Svensson, *Trans. Faraday Soc.*, **36**, 16 (1940).

26. Longsworth, L. G., and D. A. MacInnes, *Chem. Revs.*, **24**, 271 (1939).

27. Longsworth, L. G., *J. Am. Chem. Soc.*, **61**, 529 (1939).

28. Philpot, J. S. L., *Nature*, **141**, 283 (1938).

29. Svensson, H., *Kolloid-Z.*, **87**, 180 (1939);   **90**, 141 (1940).

30. Longsworth, L. G., *Ind. Eng. Chem., Anal. Ed.*, **18**, 219 (1946).

PREPARATIVE APPARATUS

31. Teorell, H., *Biochem. Z.*, **275**, 1 (1934);   **278**, 291 (1935).

32. Meyerhof, O., and W. Möhle, *Biochem. Z.*, **294**, 249 (1937).

33. Philpot, J. S. L., *Trans. Faraday Soc.*, **36,** 39 (1940).

34. Spies, J. R., H. S. Bernton, and H. Stevens, *J. Am. Chem. Soc.*, **63,** 2163 (1941).

35. Svensson, H., *Arkiv Kemi, Mineral. Geol.,* **B15,** No. 19, 1 (1942); *Chem. Centr.*, **1942, II,** 76; *The Svedberg Memorial Volume, 1884–1944,* Almqvist & Wiksells, Upsala, 1944, p. 213.

THEORY AND INTERPRETATION OF PATTERNS

36. Longsworth, L. G., and D. A. MacInnes, *J. Am. Chem. Soc.*, **62,** 705 (1940).

37. Longsworth, L. G., *J. Phys. Chem.*, **51,** 171 (1947).

38. Dole, V. P., *J. Clin. Investigations*, **23,** 708 (1944); *J. Am. Chem. Soc.*, **67,** 1119 (1945).

39. Svensson, H., *Arkiv Kemi, Mineral. Geol.*, **A17,** No. 14, 1 (1943); **B21,** No. 5, 1 (1945); **A22,** No. 10, 1 (1946).

APPLICATIONS

40. Warner, R. C., *J. Am. Chem. Soc.*, **66,** 1725 (1944).

41. Longsworth, L. G., R. K. Cannan, and D. A. MacInnes, *J. Am. Chem. Soc.*, **62,** 2580 (1940).

42. Putnam, F. W., and H. Neurath, *J. Biol. Chem.*, **159,** 195 (1945).

43. Briggs, D. R., and R. Hull, *J. Am. Chem. Soc.*, **67,** 2007 (1945); F. W. Putnam, J. O. Erickson, E. Volkin, and H. Neurath, *J. Gen. Physiol.*, **26,** 513 (1943).

ULTRASONIC VIBRATIONS

EARLE C. GREGG, JR., *Case Institute of Technology*

A. FUNDAMENTAL CONCEPTS

The term "ultrasonics" is used in acoustics to denote those sound frequencies that are beyond the upper frequency limit of the human ear. Generally speaking, ultrasonic frequencies range from about 17,000 cycles per second upward, the upper limit being determined solely by the equipment used. Recent investigations have produced frequencies as high as 500,000 kilocycles per second.

It is important to keep in mind that the laws of sound valid for the audible range are also true for ultrasonics although in the latter case other effects appear that had not been observed in the audible

range.   While these new effects are due primarily to the higher frequencies (or smaller wavelengths), many—particularly the biological, metallurgical, and chemical actions—have become evident only because of the relative ease of producing extremely large amplitudes of sound at those frequencies.   While the wavelength undoubtedly plays a role in biological reactions, a direct correlation has not yet been established, and the mass of evidence today indicates that one of the most important parameters is the sound intensity.

1. Frequency, Wavelength, and Amplitude

Whenever a sound wave travels through a given medium, the individual particles of the medium, in the simplest case, execute simple harmonic motion.   That is, each particle vibrates back and forth in a manner similar to a mass on a spring.   If the direction of vibration of the particles is in the direction of propagation of the sound, the vibration is said to be longitudinal.   If at right angles to the direction of propagation, the vibration is transverse.   Liquids and gases can support only longitudinal vibrations while solids may support both.   If the motion is simple harmonic, the maximum displacement of each particle from its rest position is called the amplitude and the number of total excursions per second is known as the frequency.

For simple harmonic motion, the time dependence of the particle displacement is said to be sinusoidal and may be represented mathematically as:

$$x = A \sin (2\pi f t) \tag{1}$$

where A is the amplitude or maximum displacement, f the frequency, t the time at which the particle is observed after having passed through a rest position, and x the displacement of the particle from its rest position.   The usual units are time in seconds, frequency in cycles per second, and amplitude in centimeters.   From the above, it also follows that the particle velocity and acceleration are:

$$v = 2\pi f A \cos (2\pi f t) \tag{2}$$

$$a = -4\pi^2 f^2 A \sin (2\pi f t) \tag{3}$$

In the discussion above, consideration was given only to the motion of each individual particle referred to its rest position.   When a sound wave traverses a medium, each particle affects the others so

that, if the first particle in a given medium is disturbed, this disturbance is transmitted through the medium by virtue of the coupling between the individual particles. In addition to this, the form or shape of the disturbance is also transmitted as long as the medium is linear (*i.e.*, obeys Hooke's law). If the disturbance at the one end is sinusoidal and if the motion of the particles is sinusoidal, it is easily seen that, for a finite velocity of propagation of this disturbance, there will be for any given particle a series of particles all with the same displacement in the same direction at regularly spaced intervals throughout the medium. The distance between two particles with the same magnitude and direction of displacement is known as the wavelength. Note here that wavelength has significance only for repeating phenomena while frequency is defined only for sinusoidal motion. If V is the velocity of propagation, λ the wavelength, and f the frequency, it follows then that $V = f\lambda$. For ultrasonic frequencies of 20 to 100,000 kilocycles per second, the corresponding

TABLE I

Sound Velocity, Acoustic Resistance (ρV), and Density (ρ) of Various Materials

Material	Temp., °C.	Density, g./cc.	Velocity, m./sec.	$\rho V \times 10^{-5}$ g./cm.^{2}/sec.
Acetone	20	0.790	1190	0.940
Benzene	20	0.879	1324	1.162
Carbon tetrachloride	25	1.595	926	1.478
Chloroform	20	1.488	1002	1.491
Ethyl alcohol	20	0.789	1168	0.920
Ethyl ether	20	0.713	1006	0.717
Glycerol	20	1.261	1923	2.430
Mercury	20	13.595	1451	19.720
Methyl alcohol	20	0.791	1121	0.888
Oil, transformer	25	0.880	1350	1.199
Oil, castor	25	0.969	1477	1.430
Water	25	0.997	1497	1.491
Heavy water	25	1.104	1399	1.543
Xylol	22	0.877	1352	1.188
Brass	20	8.56	3500	30.00
Glass	20	2.60	5500	14.30
Nickel	20	8.70	4973	43.30
Steel	20	7.83	4990	39.00
Quartz	20	2.65	5500	14.58
Aluminum	20	2.70	5104	13.80
Rubber	..	1.15	1600	1.840

wavelengths in liquids ($V \approx$ 1200 m./sec.) range from 6 to 0.0012 cm. and in solids ($V \approx$ 4000 m./sec.) from 20 to 0.004 cm.

It is important to remember that the velocity of propagation is a physical constant of the medium while the wavelength and frequency are not. For example, if a sound wave travels from one medium with a given velocity and to a second medium with another, only the wavelength will change, since the frequency is determined by the sound source and the velocity by the physical constants of the media. Other phenomena, such as reflection, occur at boundaries between media but these will be treated later. Table I contains a list of commonly used materials and their velocities of propagation. More complete lists are in Bergmann (2) and Hiedemann (3a).

2. Pressure

Since a sound wave in a medium consists of the individual particles in a regular, defined motion, it is possible to treat the wave as an alternating pressure phenomenon rather than a particle displacement as long as the physical constants of the medium are taken into account. By comparison with simple harmonic motion, we may say that the sound pressure variation with time at any point in a medium is:

$$p = P \sin (2\pi f t) \qquad (4)$$

where p is the instantaneous pressure at any time, t, and P is the pressure amplitude or maximum pressure obtained at a given point. This pressure variation may be either positive or negative, so that in a given medium two points one-half wavelength apart have a pressure differential of twice the maximum pressure. It is the large pressure and pressure differentials that account for many biological and chemical actions of ultrasound.

To convert particle displacement into pressure amplitude requires a knowledge not only of the physical constants of the medium, but also of the type of sound wave present.

3. Types of Sound Waves

In general, if a vibrating piston is a source of sound in a medium, the character of the sound wave is determined by the ratio of the wavelength of the sound in the medium to the dimensions of the piston (3, p. 108; 1, p. 147). If the wavelength is very large com-

pared to the piston, the piston acts like a point source and radiates a spherical sound wave. That is, the sound radiates in all directions with the same amplitude at any given distance from the source.

As the wavelength is diminished (source frequency increased), the sound energy tends to become concentrated more and more in a given direction (or "beamed") and for wavelengths very small compared to the piston, it approaches a plane wave. That is, nearly all the sound energy is then propagated in a unidirectional beam with cross section of about the same dimensions as the piston. The general sound pressure pattern of a source is spoken of as the "directivity" of the source. The photographs in Figure 1 of this phenomenon were made by Willard (23) utilizing the optical diffraction effects of a 10 mega-cycle ultrasonic beam. They show:

(A) Reflection of a narrow beam from glass plate G, then from upper water–air surface of the medium.

(B) Double reflection from the inside corner of a steel block (St) and subsequent absorption of the beam in a wool pad (P).

(C) Reflection of an ultrasonic beam from a cylindrical brass surface (Br) giving the familiar caustic curve.

(D) A beam focused at f by means of a planoconcave Lucite lens.

(E) An ultrasonic beam produced by the concave quartz crystal, which results in focusing.

(F) Transmission of a broad beam through a tapered aluminum plate (black in figure). Plate transmits for thicknesses that are multiples of $\lambda/2$.

(G) Diffraction of beam around a wire 21 λ in diameter; note that sound reappears in the center of the shadow.

It is readily seen that, if thermal and viscous losses are not present, the pressure or amplitude of a spherical wave diminishes inversely as the distance from the source while there should be no variations at all with distance for a plane wave in a homogeneous medium.

From the wavelengths calculated previously for ultrasonic frequencies and considering the usual physical size of a sound source, it is obvious that most ultrasonic waves are plane waves and may be treated as such.

The problem of what happens to a sound wave when it strikes a boundary or object is also decided by the ratio of the wavelength to the dimensions of the object (2, p. 299). If the wavelength is small compared to the dimensions of the object (about $\frac{1}{10}$ the object size or smaller), a reflection takes place that may be treated in the same
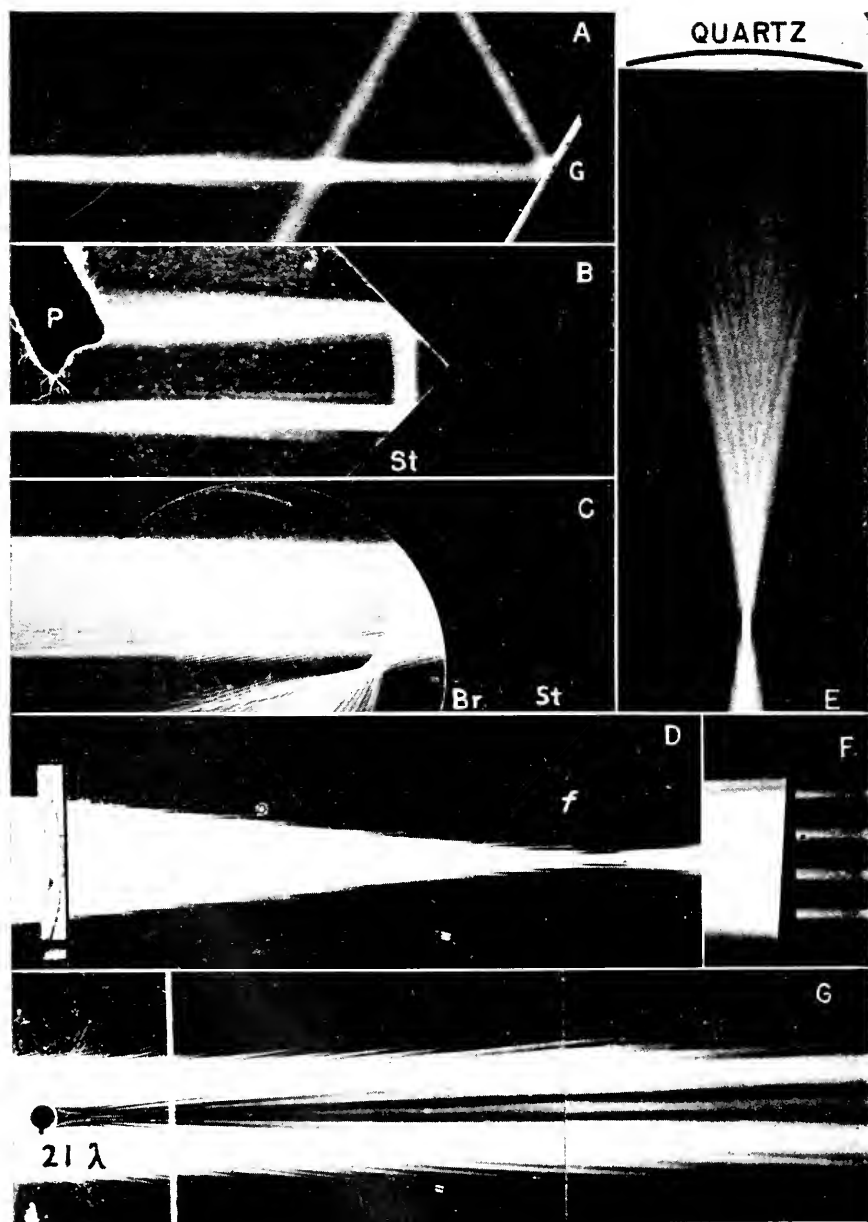
Fig. 1. Ultrasonic paths in water, photographed by virtue of their optical effect on the medium. The ultrasonic beam (frequency 10 megacycles per second) enters from the left in each figure except E. (Courtesy *Bell Telephone Laboratories*.)

manner as light reflections in optics. Likewise the portion of the
sound energy transmitted through the object may also be treated in a
manner similar to light incident on lenses and prisms since the object
undoubtedly has a different sound velocity than the medium in
which it is immersed (and hence an "acoustic" refractive index).
Such sound lenses and prisms have been constructed and constitute a
very important tool in ultrasonics. Photograph D in Figure 1 con-
stitutes an experimental verification of this made by Willard (23).

If the wavelength of the sound is large compared to the object
(about twice object size or larger), the object both refracts and reflects
the beam and creates a new type sound field in its vicinity. The
magnitude of the effect is determined by the ratio of wavelength to
object size. The larger this ratio, the less the effect until finally the
object becomes "invisible" to the sound. For intermediate ratios,
the problem is very difficult to calculate.

For plane waves, it is easy to visualize a condition in which the
sound is reflected from a plane boundary back to the source and then
back to the boundary again. If the various direct and reflected
waves present at one point all vary in the same manner at the same
time (that is, if the reflector is an integral number of half wavelengths
away from the source), a "stationary" wave is said to be present. A
stationary wave is characterized by the fact that, as one moves
through the sound path, the amplitude of vibration passes through
well defined maxima and minima. These are known as loops and
nodes when dealing with dust patterns, organ pipes, and vibrating
strings in the audible sound range.

Even when dealing with plane ultrasonic waves, the container (or
boundaries) has a very important bearing on the type of sound field
present owing to these multiple reflections. The results of any given
investigation that did not take this into account might be difficult
to interpret correctly. (For a more complete discussion, refer to
Section C1.)

4. Intensity

Sound intensity is defined as the energy that passes through a
unit area in unit time. The dimensions are usually ergs per second
per square centimeter, or watts per square centimeter. For a plane
sound wave, the intensity, I, may be shown to be (1, p. 55):

$$I = (\rho V/2)(2\pi f A)^2 = (P^2/2\rho V) \tag{5}$$

where ρ is the density of the medium and the other symbols are defined as before.   This relation allows the pressure to be calculated in terms of the amplitude:

$$P = 2\pi f \rho V A \qquad (6)$$

For a source of area S (S large compared to λ^2), the mean power radiated becomes:

$$\text{Power} = \tfrac{1}{2}\rho V (2\pi f A)^2 S \qquad (7)$$

For example, consider a typical quartz disc vibrating at a frequency of 500 kilocycles per second and delivering a sound intensity of about 10 watts per square centimeter (10^8 ergs/sec./cm.^{2}) into water.   For water, $V = 1.48 \times 10^5$ cm. per second and $\rho = 1$, so that we have from equation (5):

$$P = (2\rho V I)^{1/2} = 5.4 \times 10^6 \text{ dynes/cm.}^2$$

or about 5.4 atmospheres.

In this example then, the pressure alternates from $+5.4$ atmospheres to -5.4 atmospheres 500,000 times a second.   To calculate the amplitude of vibration of the water molecules, we find from equation (6) that:

$$A = (P/2\pi f \rho V) = 1.16 \times 10^{-5} \text{ cm.} \qquad (8)$$

From the laws of simple harmonic motion (see equation 3), we further find that the maximum acceleration of the particles is:

$$a = 4\pi^2 f^2 A = 1.14 \times 10^8 \text{ cm./sec.}^2$$

or an acceleration about 10^5 greater than that due to gravity.   These large accelerations account for much of the coagulating action of ultrasonics and other related effects.   Considering the above values, ultrasonic fields have been correctly described as "all acceleration and no motion."

It is worth while to point out here that, since the density of the medium appears in the intensity equation, very large amplitudes of vibration are required to produce a given acoustic intensity in a gas as compared to a liquid, or solid.   This fact coupled with energy conversion efficiencies means that far different generating devices must be used in the various media.   The magnetostrictive and piezoelectric generators discussed later are adapted primarily to production of

ultrasound in liquids and solids while loudspeakers and sirens are used for airborne ultrasound.

5. Reflection and Transmission at Boundaries

Whenever a plane sound wave strikes a boundary between two media, part of the energy is reflected to the first medium and part is transmitted through the second. If ρ_1 and V_1 are the density and velocity of propagation of the first medium and ρ_2 and V_2 those of the second, then, if the wavelength of the sound is small compared to the dimensions of the boundary, we have for the ratio of reflected to incident energy (*1*, p. 271):

$$R = \left(\frac{\rho_1 V_1 - \rho_2 V_2}{\rho_1 V_1 + \rho_2 V_2}\right)^2 \tag{9}$$

This gives, for the percentage of sound energy reflected from the boundary, water–glass 66%, oil–glass 70%, water–steel 85%, and oil–steel 88%. For sound energy incident on an air–water interface or vice versa, the above relation indicates that 99.9% would be reflected.

Consider now a plate of thickness d, density ρ_2 and velocity V_2 immersed in a liquid of density ρ_1 and velocity V_1. If a sound wave present in the liquid strikes the plate at right angles, the ratio of reflected to transmitted energy becomes:

$$R = \left(\frac{\rho_1 V_1}{\rho_2 V_2} - \frac{\rho_2 V_2}{\rho_1 V_1}\right)^2 \bigg/ \left[4 \cot^2 \frac{2\pi d}{\lambda_2} + \left(\frac{\rho_1 V_1}{\rho_2 V_2} \frac{\rho_2 V_2}{\rho_1 V_1}\right)^2\right] \tag{10}$$

where λ_2 is the wavelength of the sound in the plate. It is easily seen that, if the plate is infinitely thick, the above formula reduces to the one previously given for two media. On the other hand, if d is small compared to λ_2, virtually all the sound energy is transmitted through the plate. Note also that, if the product ρV for the plate is the same as that for the medium, complete transmission results. This criterion has been used in many instances in the selection of materials for underwater sound equipment.

For intermediate thicknesses almost complete transmission will result for $d = n\lambda_2/2$, where n is an integer. Practically speaking, n must be confined to small numbers (1 through about 10). This fact is extremely important in biological work in which the ultrasonic energy must traverse the glass wall of a beaker or flask in order to

reach the sample being treated. Figure 2 shows the relationship between sound transmission and the ratio of thickness to sound wavelength for a glass plate immersed in water. Since V_2 for glass is on the order of 5500 m. per second, the wavelength of sound in glass at a frequency of 1000 kilocycles is 0.55 cm. While Figure 2 shows that for this condition a thin-bottomed flask will suffice for a relatively large transmission of sound energy, the effect cannot be overlooked at higher frequencies.

Figure 1F is a photograph by Willard (23) showing the selective thickness transmission of 10 megacycle ultrasonics by an aluminum wedge.
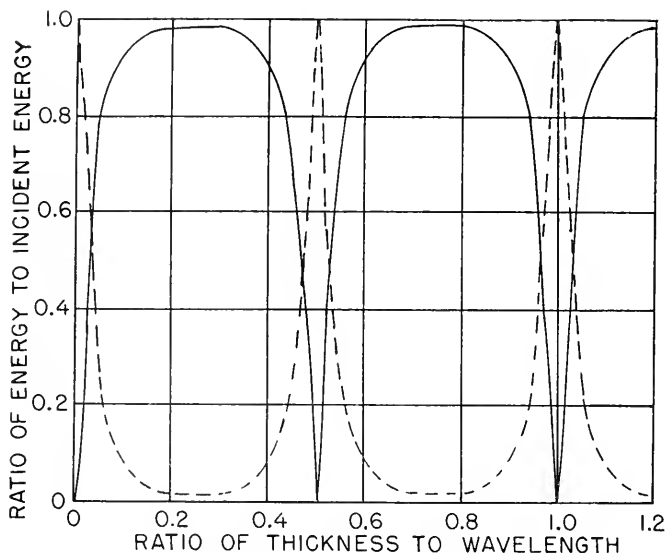
Fig. 2. Behavior of plane sound waves incident normally on a glass plate of thickness d: (—) reflected energy; (- - -) transmitted energy.

When sound waves in general meet a boundary, they exert upon it, in addition to the alternating pressure mentioned, a steady unidirectional pressure. This steady pressure is called the radiation pressure and is responsible in the case of a liquid–gas boundary for the actual distortion of the liquid surface. For very intense sound waves in a small liquid volume, as is usually encountered in chemical and biological investigations, particles of the liquid may actually be ejected as far as 10 to 30 cm. above the surface.

6. Absorption and Scattering

Whenever a sound wave traverses a medium, there is naturally some loss of energy to the particles of the medium. While these losses may be separated into various types, it will suffice here to consider all losses together. These losses will generally appear as heat and raise the temperature of the medium.

In all cases, the sound intensity, I_0, of a plane sound wave decreases in a liquid by passage over a distance d to the value

$$I = I_0 e^{-2\alpha d} \tag{11}$$

where α is the over-all amplitude absorption coefficient of the medium. Theory has shown that the absorption coefficient varies directly as the square of the frequency. While some deviations from this law have been found, it is the general practice to list the value of α/f^2 that is more a characteristic of the material than α. A list of values of α/f^2 for some common materials is shown in Table II. These values may be assumed to hold approximately over a fairly large range of frequencies. The absorption coefficients of many more materials and mixtures have been investigated and listed by Willard (23).

TABLE II

Sound Amplitude Absorption Coefficients of Various Liquids

Material	$(\alpha/f^2) \times 10^{16}$ cm.$^{-1}$ sec.^{2}
Carbon disulfide	74
Glycerol	26
Benzene	8.3
Carbon tetrachloride	5.7
Chloroform	3.8
Kerosene	1.1
Toluol	0.9
Acetone	0.64
Water (distilled)	0.33

To calculate the distance over which half the energy is lost, we merely place $I = \frac{1}{2}I_0$ and then $2\alpha d_{1/2} = \ln 2 = 0.693$ or:

$$d_{1/2} = 0.693/2\alpha = 0.347/\alpha$$

In the case of benzene at 950 kilocycles, for example, $\alpha/f^2 = 8.3 \times 10^{-15}$, $\alpha = 0.0075$ cm.$^{-1}$ and so $d_{1/2} = 46$ cm. In other words, the

sound wave would have to travel 46 cm. before its intensity would be reduced to one-half its initial value.

It is important to note that the coefficients in Table II are for homogeneous isotropic materials.   The presence of foreign matter will increase the absorption coefficient not only by absorption of energy by the foreign matter itself, but also by scattering the sound energy out of the beam.   This effect of foreign matter (particularly dissolved gases) on the absorption of energy is of extreme importance in biological work.   More will be mentioned about this later.

7. Cavitation and Degassing

Whenever a sound wave traverses a liquid in which there are dissolved gases, small groups of gas bubbles are formed.   For sound of low energy, it has been found that these bubbles are caused by the union of microscopically small bubbles as they move toward the nodes of a stationary wave.   For larger energies, the negative pressures involved actually cause the emergence of the gas dissolved in the liquid and greatly increase the rate of bubble formation.   Even with no dissolved gases, the large negative pressures and hence large stresses in the liquid will cause small hollows or cavities to be formed, which then become filled with the vapor of the surrounding liquid. These hollows of course disappear when the sound beam is turned off. The formation of these hollows by the literal tearing apart of the liquid is known as cavitation.

If there are dissolved gases, the local action caused by the gas filling these cavities is tremendous and explains in part most of the observed biological and chemical actions of ultrasound.   There is not only tremendous local agitation but also high local temperatures and possibly electrical potentials due to the frictional losses involved as the gases escape into the cavities.   The absorption coefficient under these circumstances is very large.   Investigation has shown that the power per square centimeter required to produce cavitation (and hence bubble formation) depends mostly on the external pressure while the energy required to produce a given volume of gas depends on both the external pressure and frequency.

In general, cavitation occurs in a light liquid filled with air when the sound pressure is on the order of the hydrostatic pressure at the point in question plus the external pressure on the liquid.   For example, for water at atmospheric pressure and negligible depth, the acoustic intensity required to produce cavitation is:

$$I = P^2/2\rho V = 0.34 \times 10^7 \text{ ergs/sec./cm.}^2 = 0.34 \text{ watt/cm.}^2$$

When liquids are degassed, their natural cohesive pressure becomes effective and they will stand a negative acoustic pressure. It has been found that under these circumstances the total negative pressure required to cause cavitation is equal to the sum of the cohesive pressure (tensile strength) and the ambient pressure. The cohesive pressure appears to be a variable quantity and depends to quite an extent on the previous history of the liquid. Once a liquid has cavitated, it will cavitate at a lower acoustic pressure. It will also require some time to return to its previous state.

It has been found that the optimum pressure for pronounced cavitation in water is approximately two atmospheres. This corresponds to an intensity of 1.35 watts/cm.^{2} Briggs, Johnson, and Mason (22) have attained higher intensities than this without cavitation by pulsing or driving the source for very short time intervals. There is seemingly some time delay in producing cavitation. Briggs, Johnson, and Mason have formulated a theory of this time delay based on Eyring's theory of viscosity, plasticity, and diffusion, which agrees quite well with experiment. Harvey (32) and collaborators have treated in a similar manner the formation of gas bubbles in blood and other liquids. Among other interesting phenomena, Harvey has found that subjecting liquids to a high hydrostatic pressure (1000 atmospheres) prior to investigation results in a condition in which only very severe blows will cause bubbles to form even when the container is exhausted to the vapor pressure of water. See also Novotny (36) and Pease and Blinks (37).

For heavy viscous liquids, the power required for cavitation is approximately two to four times that required for light liquids. This is explained by the fact that viscous liquids have a high cohesive pressure. A linear relationship apparently exists between the sound cavitation amplitude required and the viscosity of the medium. As far as degassing is concerned, to obtain 1 cc. of air per second from water saturated with air, Sorensen found that it required 51.2 kilowatts at 194 kilocycles, 72.6 kilowatts at 380 kilocycles, and 87.4 kilowatts at 530 kilocycles, a rather inefficient means of degassing a liquid.

From all the above figures, it is seen that, in order to produce cavitation and the ensuing biological actions, relatively large amounts of ultrasonic power are required. Considering the fact that the loudspeaker in an average radio radiates about 10^{-8} watts per square

centimeter, coupled with the density effect of the media mentioned
previously, it is obvious that other techniques and sources must be
employed than those usually encountered in air acoustics.

B. PRODUCTION OF LARGE AMPLITUDE ULTRASOUND IN LIQUIDS

1. Magnetostriction Devices

One of the better methods for the production of large power ultra-
sound in liquids involves the use of magnetostrictive materials.
The phenomenon of magnetostriction is the change in length of a rod
or tube of ferromagnetic material when it is introduced into a mag-
netic field parallel to its length.   This change in length, like most mag-
netic phenomena, is reversible.   Hence if an initially unmagnetized
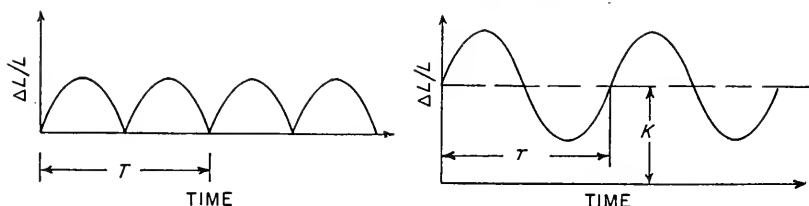
Fig. 3.   Relative change in length of a magnetostrictive rod as a function of
time: at left, without a biasing d.c. field; at right, with a biasing field. T is
the period $1/f$ of the alternating magnetic field. K is the relative stretching of
the rod produced by the biasing field.

rod is brought into an alternating magnetic field, it will contract and
expand with twice the frequency of the field.   On the other hand, if
the rod is initially magnetized, it will vibrate with the same fre-
quency as the field.   This is illustrated in Figure 3.   If the natural
period (period $= 1/f$) of the premagnetized rod is the same as that
of the alternating magnetic field, the amplitude of vibration will be a
maximum and since the vibrations of the rod are longitudinal, sound
waves will be emitted from the ends of the rod.

Figure 4 shows typical curves of magnetic field versus relative
change of length for a few magnetostrictive materials.   While nickel
shows a relatively large effect and is recommended for ultrasonic
sources, other materials such as Permalloy and Invar have been used
(5).   The relative change of length as plotted is for a free bar of the

material. The actual relative change in length obtained in operation
is determined not only by the material used but also by the method of
clamping the rod and by the acoustic load presented to the rod. Ex-
amination of the graph shows that for nickel the addition of a pre-
magnetizing field on the order of 50 oersteds will not only prevent
frequency doubling, but will also place the operating point of the
rod at a position where the relative change in length is fairly large
for a given change in magnetic field. This point is rather critical
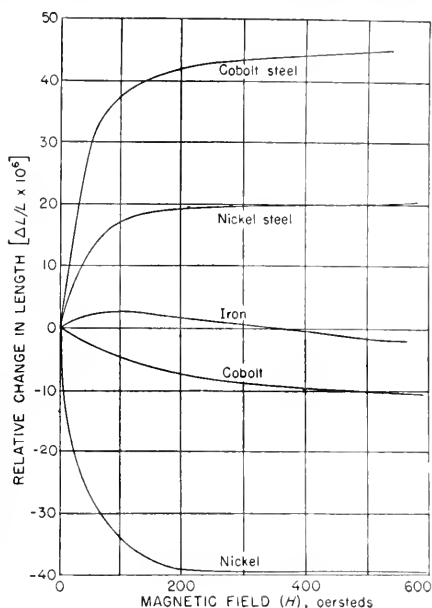
Fig. 4. Relative change in length
as function of magnetic field
strength for various magnetostric-
tive materials. Negative values
signify contraction; positive values
denote expansion with increasing
magnetic field.

for best operation of the rod. As a general rule, the rods are not
premagnetized to this point but rather placed in a special magnetic
yoke that will allow the superposition of both d.c. and a.c. fields.
It is important to note that the peak a.c. field cannot exceed the bias-
ing d.c. magnetic field without producing excessive distortion of the
sound emitted by the rod. Consider as a simple example a nickel
rod 10.55 cm. in length vibrating at 20 kilocycles with a biasing field
of 50 oersteds. If we assume that the superposed a.c. field has an
amplitude of 10 oersteds, this will produce, as seen in Figure 4, an
amplitude of vibration of the free bar of:

$$A = (5 \times 10^{-6})(10.55) = 0.53 \times 10^{-4} \text{ cm.}$$

If we assume that the effect of clamping and a water load are negligible, it follows then from the previous relationships that this amplitude in water represents an intensity of 0.47 watt per square centimeter.   By using higher magnetizing fields and operating the bar at resonance (24 kilocycles), some investigators have achieved intensities as high as 20 watts per square centimeter.

As is well known, the frequency of vibration of a rod depends upon where it is clamped.   Although there are many different modes of vibration, the rod will vibrate strongly only when it is clamped at a node, the position of the desired node along the rod being determined by the harmonic selected.   However, since the energy output diminishes considerably when a rod is clamped to produce higher harmonics, it is the general practice to clamp rods in the middle in order to produce intense vibration at the fundamental frequency.   This central clamp also provides a suitable mechanical support for the rod. In practice, the position of this clamp may have to be adjusted for maximum output since the two ends of a rod are often loaded unequally;   as a result a node will not occur at the exact center.   Although magnetostrictive sources of ultrasound have been developed that utilize rods or tubes clamped at one end, these have not been generally used or found adaptable to biological research.   Tubes clamped in the middle have proved the simplest to construct and operate.

The natural frequency of any rod or tube is given by:

$$f_n = (n/2b)(E/\rho)^{1/2} \qquad (12)$$

where E is Young's modulus, ρ the density, b the length, and n the harmonic for which it is clamped (1, p. 135).   Since the frequency is inversely proportional to the length, one limit to the frequencies attainable is the physical size of the rod.   For example, if a nickel rod were to vibrate at a fundamental frequency of 50 kilocycles, its length would be only 5.07 cm.   Rods shorter than this become not only less convenient to handle but also more difficult to excite. Harmonics may be used, but, as mentioned, there is a resultant loss of intensity.   Pierce (5) has designed some special rods for high frequencies.

When any metal is introduced into an alternating magnetic field, localized (eddy) currents are generated in the metal and thus part of the electrical energy is converted to heat.   Hence, the presence of these parasitic currents lowers the efficiency of conversion of electrical

to sound energy.   Since eddy currents are generated in a plane nor-
mal to the direction of the magnetic field and their magnitude is
proportional to the square of the width of the conducting metal in
this plane, it is common practice in magnetostriction devices to use
thin-walled tubes instead of rods and to slit these tubes in the direc-
tion of the magnetic field.   The tube need be split only over the
length that is in the magnetic field.   The end of the hollow tube that
enters the sound medium is generally sealed over with a flat plate or
plug to increase the radiating surface and to provide a liquid seal
for the container.
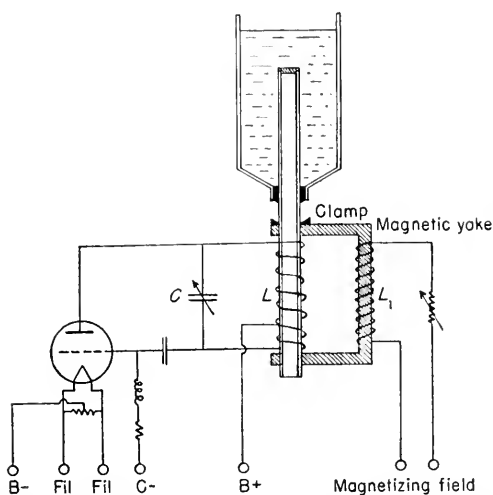
Fig. 5.   Schematic sectional view of a typical magnetostriction oscillator.
B−, B+, Fil, and C− are the electrical supply voltages.

A schematic diagram of a typical magnetostriction oscillator is
shown in Figure 5.   While more complex circuits have been used,
most biological or chemical investigations do not require them as long
as the required power and proper measuring devices are available.
The frequency of oscillation of the simple electron tube oscillator
pictured is determined by the relationship $f = 1/[2\pi(LC)^{1/2}]$ where
L is the inductance of the coil shown with the magnetostrictive tube
present and C is the capacitance.   By varying C the frequency of the
oscillator can be adjusted to the natural period of the rod, at which
point the largest amplitude of vibration will be produced. L_1 is the
d.c. coil necessary to produce the proper static operating field for the

tube. At high frequencies, the magnetic yoke shown must be of laminated electrical steel to minimize eddy current losses. In some cases, L and L_1 are wound on the same yoke. The power developed by most oscillators for these frequencies ranges from 500 to about 2000 watts. At these higher powers, care must be taken not to overheat the nickel tube because of the eddy currents present even though the tube is slit. Water cooling has been used in some cases.

While it is possible to produce sound at frequencies other than the natural period of the metal tube, the amplitude of vibration is extremely low and recourse is generally made to tubes of different lengths. If the frequency range desired is too large, different a.c. coils should also be used since the efficiency of the electron tube oscillator at various frequencies depends upon the coil construction. Salisbury and Porter (6) describe a very excellent magnetostriction generator of this type adapted to chemical and biological investigations. A 9 kilocycle commercial unit is available (40). Magnetostriction generators are used primarily because of their ruggedness, simplicity, and ease of producing high power. Their main disadvantages are the limited frequency range, the great dependence of frequency on tube temperature, and the breadth of the natural resonance curve. The last factor results from the change in elastic constants of the metal tube with magnetization. The effect of clamping and acoustic loading on the resonance curve will be discussed later. To produce high sound energies at frequencies above 50 kilocycles, recourse is generally made to piezoelectric crystals, but a 100 kilocycle magnetostriction apparatus has recently been described (3d).

2. Crystal Apparatus

Ultrasonic vibrations may also be generated by the production of mechanical strains in certain crystals when electrical charges are placed on the proper crystal faces (4). From the viewpoint of electrical, mechanical, and chemical properties, the best crystal today is quartz, although tourmaline, Rochelle salt, and ammonium dihydrogen phosphate have been used. The two latter crystals, however, because of thermal and cavitation effects at high amplitudes, are not generally useful in chemical or biological work. Tourmaline would be satisfactory except that it is not available in large crystals.

There are many ways to cut a quartz crystal and still have it active electrically. However, most of these cuts have special properties that are of little advantage in producing high energy sound.

One of the best cuts for this purpose is the x-cut crystal illustrated in Figure 6. The faces normal to the x axis are those to which the alternating electrical field is applied and the resultant motion of the crystal is one of alternate expansion and contraction between the two faces. This means that, if one face of the crystal were cemented to a solid block, the other face will move up and down with a pistonlike action. While the amplitudes of vibration are generally small for
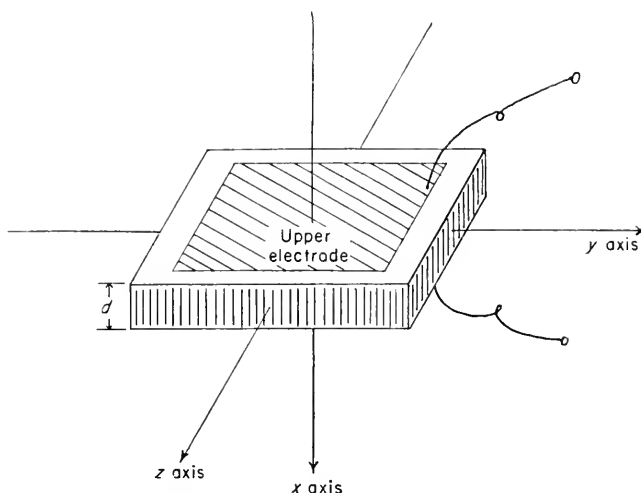
Fig. 6. Arrangement of crystal axes in an X-cut quartz plate. An alternating potential applied to the electrode surfaces will cause contraction and expansion in thickness d.

the reciprocal piezoelectric effect, the power delivered can be large owing to the high frequencies involved.

For an x-cut crystal, the type of vibration set up by an alternating electric field is similar to that produced in a magnetostrictive rod. That is, the crystal has a natural period of vibration determined by its elastic constants and its thickness (in the x direction). The fundamental frequency of vibration is:

$$f = (c/\rho)^{1/2}/2d \tag{13}$$

where, for x-cut quartz, $c = 85.46 \times 10^{10}$ dynes per square centimeter, $\rho = 2.65$ g. per cubic centimeter, and d is the thickness in centimeters

along the x axis.   For very low frequencies, some x-cut quartzes are cut in the form of rods;   however, magnetostriction is superior at those frequencies.

The efficiency of generation of ultrasound depends not only on the purity of the quartz but also on the manner in which it is clamped and the type of electrodes on the faces.   As a general rule, the faces are metallized by evaporation of copper or silver and then thickened by electroplating.   Some investigators have merely laid thin metal sheets or gauze on the faces but this is not as satisfactory as a metallic plating.   Care must be taken not to have the electrodes too near the edge of the crystal since the dielectric breakdown strength diminishes near the edge.   Once a breakdown occurs, a conduction path results and the crystal usually cracks from the heat generated.

At a given frequency the amplitude of vibration of a piezo crystal is directly proportional to the voltage applied to its faces.   For large vibrations then, rather large alternating voltages must be applied to the crystal.   Care must be taken not to puncture the crystal nor to mechanically shatter it by forcing it at too large an amplitude.

Some investigators have used potentials as high as 34,000 v. but this has been for short time service with painstaking insulation precautions.   The potentials usually range from 1000 to 10,000 v. depending on the thickness of the crystal involved.   About 2000 v. per millimeter is a satisfactory field strength.   A special crystal cut—Straubel contour (2, p. 24)—has been developed which produces a more uniform motion of the crystal faces.   This in turn allows higher voltages and hence higher amplitudes of vibration with less danger of fracture.

The type of mounting and acoustic loading of the crystal play an important part in the design and use of a crystal generator.   Several different arrangements are shown in Figure 7.   In a, the crystal and the upper thin metallic electrode are merely clamped by a ring to a metallic block.   This block should be of lead since lead is less likely to shatter the crystal than a harder metal.   The most favorable size of the upper clamping electrode and clamping electrode pressure must be found by experience.   Electrodes with too small a hole produce small ultrasonic outputs while large holes lead to irregular excitation of the quartz.   If the crystal is plated, the latter factor is of little consequence and small clamping pressure need be used.   In b, the crystal is backed by an air space rather than a metallic block. This means that, since little sound energy is radiated into air, most

of the energy that previously went into the mounting will now be reflected back and out the top face of the crystal.

As a general rule, most quartz crystals are driven under oil because of its electrical insulating properties. If any other material is to be irradiated, it is usually introduced into the sound beam in the oil by means of a thin-bottomed beaker as shown. Salisbury and Porter (6) have estimated that about one-third the energy radiated into the oil bath can be transmitted to a liquid inside the glass vessel. In case the crystal becomes too warm under continuous operation,
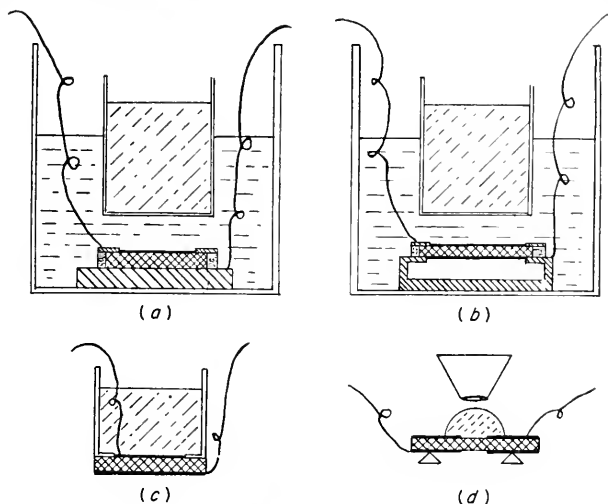
Fig. 7. Quartz crystal mountings as used in chemical and biological research.

the oil may be circulated and cooled by external means. In c, a bottomless beaker is cemented to the crystal and the one crystal face drives directly the material in the beaker. Connection can be made directly to the upper face of the crystal (if it is plated) or to the liquid if it is a conductor. In this manner, intermediate media and beakers are avoided; however, the insulation and breakdown problems remain.

In d, the material under observation is placed on a crystal and observed through a microscope. The crystal plating may be opened at the point of observation to permit the passage of light if desired. In this case, the crystal cannot be driven too hard because of possible shattering and no acoustic measurements of the sound field can be made. Without the latter, the observations become very empirical.

Figure 8 shows a hand unit designed by Pohlmann (30) for therapeutic purposes. The membrane is pressed against the flesh of the subject.

As mentioned previously, the type of clamping and acoustic load determine to a great extent the amplitude of vibration of either a
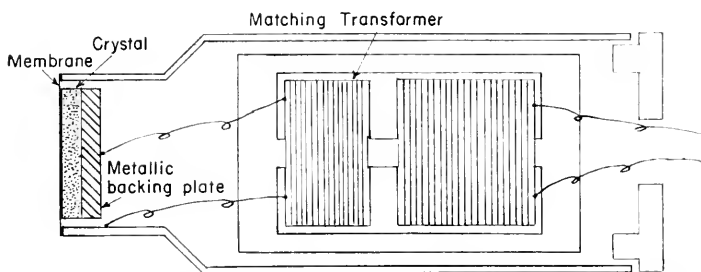
Fig. 8.   Hand ultrasonic unit designed for therapeutic purposes.
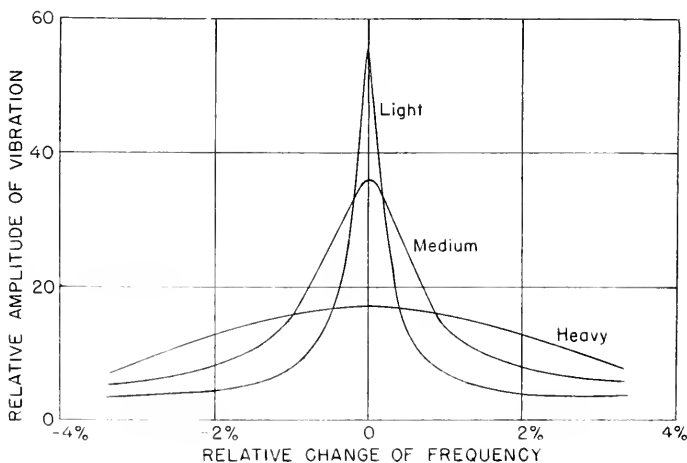
Fig. 9.   Effect of light, medium, and heavy damping on the resonance curve of a vibrating quartz crystal.

quartz crystal or magnetostrictive rod. Figure 9 shows the effect of various loads or "damping" on the resonance curve of a quartz crystal. The "medium" damping condition corresponds approximately to the effect of a liquid load on a vibrating quartz crystal. Crystal clamping devices damp the crystal vibration in a similar manner. Some investigators have taken pains to support a crystal in the me-

dian plane but this is largely wasted effort unless the crystal radiates to the same extent in both directions. This is rarely the case and when the two faces see different loads, the median plane is no longer a nodal plane.

Consider a crystal (or tube) radiating sound into a medium of infinite extent. Under these circumstances, the medium offers some resistance to the motion of the crystal (it must if power is to be delivered to the medium) and effectively lowers its amplitude as shown before. Now introduce a reflecting boundary in front of the radiating surface. The waves reflected to the face of the crystal then act either to damp the motion or to augment it depending on the phase of the reflected wave relative to the motion of the vibrating surface. As is readily seen, the relative phase of the two is determined by the distance between the generating and reflecting surfaces. If the distance is adjusted so that the two are in phase (an integral number of half wavelengths) a standing wave pattern results and the power delivered by the source is a maximum. This acoustic loading of the crystal is reflected back to the electrical circuit and actually causes the driving currents or voltages to change as the loading is changed. Use is made of this phenomenon in acoustic interferometry (2, p. 57).

While the sound fields used in most biological work are too complex to be analyzed in terms of simple plane wave reflections between two surfaces, it remains that the introduction of specimens into the sound field not only changes considerably the character of the sound field but also changes the amplitude of vibration of the source. This should be remembered when measurements of the sound field are attempted.

To obtain a variable frequency source of sound it is possible to excite a quartz crystal to any amplitude at any frequency however far from resonance as long as the exciting voltage does not exceed the dielectric breakdown strength of the crystal. As far as mechanical shatter is concerned, it may be said that the danger of shatter with forced vibrations at high voltage is no greater than resonant vibrations at low voltage. However, the forcing of a crystal off resonance is very uneconomical as is seen from the resonance curves. Even if it were economical to use high voltage at off-resonant frequencies, the crystal would most likely be shattered, even with liquid damping, if by chance the frequency happened to pass through resonance. Resonant frequencies and a series of crystals are usually employed

to cover a given frequency range.   It should be kept in mind, however, that a crystal vibrating with a reasonable amplitude at resonance in a liquid will certainly shatter if driven at the same voltage in air.

For producing very high sound intensities at a point, Greutzmacher (7) developed a special cut of quartz crystal with a concave face.   In this manner, the energy was concentrated at a focal point resulting in an increase of as much as 150 times the energy of a small surface element close to the crystal.   This type crystal cut obviously allows the production of high sound intensities at relatively low voltages with little danger of shattering the crystal.   Using a crystal of this nature, at frequencies from 638 to 1000 kilocycles, Tumanski (8) produced jets of oil 70 cm. high, projected above the free surface of the liquid.   Figure 1E is a photograph by Willard (23) of the sound field produced by such a concave ultrasonic crystal.   Other investigators have used combinations of crystals all aimed toward a focal point, but under these circumstances care must be taken to assure the correct phasing of the sound waves at the focal point.

As was mentioned previously, it is possible to construct acoustic lenses that will focus sound just like glass lenses in optics.   While this has not yet been applied to biological research, it has proved remarkably useful in other phases of ultrasonics (27).

As long as the wavelength of the sound is small compared to the dimensions of the lens, the lens laws of optics apply to ultrasonics (for example, the "thick lens" relationship).   However, it is important to remember that if the sound velocity in the lens material is higher than that of the surrounding medium (as is usually the case) the relative refractive index is less than unity and an acoustic lens shaped like an optical diverging lens will actually converge the sound. In ultrasonics, lens materials such as polystyrene or carbon tetrachloride in smooth aluminum containers (thickness and imperfections small compared to λ) have been used.   In general, such devices should prove much more economical than special cuts of quartz crystals if their losses are small.   Figure 1D shows the sound field produced by such a lens.

Figure 10 shows two methods of exciting quartz crystals.   Both are simple electronic oscillators of the Hartley type and differ only in the manner of coupling to the crystal.   In a the crystal is connected directly across the tank capacity, which is a source of high alternating voltages.   The disadvantages of this scheme are that the voltage at-

tainable is definitely limited by the circuit and also that the crystal is at a high d.c. potential. The latter factor complicates the handling of the crystal and mount. In b the tank inductance forms the primary of a high voltage transformer, the secondary of which is connected to the crystal. Very high voltages are attainable if the crystal capacitance, stray capacitance of the secondary, and the secondary inductance form a circuit resonant at the natural period of the quartz. As with the magnetostriction oscillator C is varied until the electronic circuit oscillates at the crystal resonant frequency. In b, L_2 may be separated from L by a fairly large distance and connected to it by a
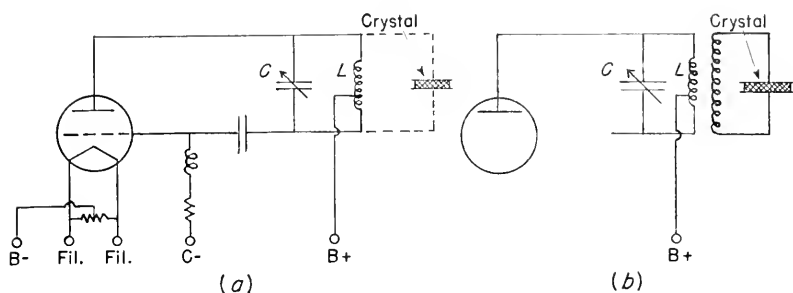
Fig. 10. Typical electronic oscillator with two methods of coupling to a piezoelectric crystal: (a) direct-coupled; (b) transformer-coupled. B−, B+, Fil, and C are the electrical supply voltages.

low impedance feeder link. In this manner, the high voltage source and crystal may be mounted in an oil-filled tank away from the electronic oscillator and greatly enhance the flexibility of the system. Other coupling systems have been described in the literature (29).

As with most electronic circuits, there are a large number of variations, each with different characteristics. The choice remains with the experimenter as to the type used since this is usually decided by the material and funds available. Salisbury and Porter (6) and Smith and Stumpf (29) describe some typical circuits and over-all generator designs used especially in chemical and biological studies. Present day sources of commercial ultrasonic equipment are listed at the end of the bibliography. Current books contain design data (3b,c).

While the power outputs of most generators of this type vary from about 200 to 2000 watts, the maximum acoustic power delivered by the crystal depends on mounting, frequency, and acoustic loading,

Wood and Loomis have recorded a value as high as 35 watts per square centimeter but this is an unusual case. 10 watts per square centimeter is the usual limit although even here there is danger of fracture. As a general rule, the sound radiation pressures are sufficient to produce fountains of oil 3 to 20 cm. above the free surface of the liquid.

Crystal efficiencies range from 75 to 90%. This means that 75 to 90% of the electrical energy supplied to the crystal is actually converted into sound energy. On the other hand, the over-all efficiency of the system including oscillator and crystal runs about 20 to 40%. Efficiencies of magnetostriction sources are about the same order of magnitude.

3. Sound Field Measurements

As with any research, a knowledge of the magnitude of the factors involved is of paramount importance. Some of the discrepancies in the interpretation of biological action of ultrasound today is a direct result of not recording or measuring the character of the sound field used. To measure the electrical quantities alone is not sufficient. The three most important parameters in sound fields of the type used in biological and chemical research are the frequency, amplitude (either displacement or pressure), and intensity. A measure of any two allows the calculation of the third.

The frequency is most simply determined by measuring the frequency of the driving oscillator with a standard radio wavemeter. The amplitude presents, on the other hand, a very difficult problem. Salisbury and Porter (6) developed a special circuit that measures the amplitude of vibration of a magnetostrictive tube. This method consists simply of arranging the tube to form one plate of a two plate air-spaced condenser and measuring the alternating voltage produced when the distance between the two plates is changed in a sinusoidal motion (i.e., vibration of the tube). An instrument of this type can be calibrated on an absolute scale. It will only measure, however, the amplitude of the source and not that at any given point in the sound field. To measure the amplitude at any point in the sound field, recourse must be made to devices that convert the sound energy to some other measurable form of energy (1, p. 373; 2, p. 39; 3, p. 28; 27). A simple and accurate instrument for this is again a piezoelectric crystal, but in this case using the converse of the effect discussed pre-

viously, that is, the production of charges on certain faces when the crystal is mechanically strained. Rochelle salt, ammonium dihydrogen phosphate, quartz, and tourmaline have all been used for this purpose.

While care must be taken to see that the crystal and holder are small enough so as not to interfere with the sound field, in most biological investigations the whole liquid volume is filled with sound energy and the effect of a small pick-up will be negligible. Rochelle salt has been used considerably for low sound energy detection while for high sound levels, quartz and tourmaline have proved the most useful. Such units may be used for relative sound measurements at one frequency or calibrated and used on an absolute basis (27). Magnetostriction pick-ups have also been devised for this purpose (9). Other sound receivers based on thermal and pressure effects have been used but have not proved as useful or accurate as the piezoelectric devices mentioned above.

In some cases, it is desirable to know the total energy delivered to a given volume. This is done most simply by measuring the temperature rise for a given exposure time and calculating the heat energy delivered to the volume. The usual precautions should be taken to insure that no heat is lost by radiation or conduction. This heat energy will then represent the total acoustic energy delivered by the crystal since only a negligible amount of acoustic energy is delivered to the surrounding air.

As a simple example assume that 500 ml. of oil of specific heat 0.5 and density 0.8 is raised 2°C. in one minute by a 2 cm. diameter quartz disc vibrating at 500 kilocycles. Assuming that all the energy of the quartz is delivered to the oil, we see that the total energy added is 400 cal. or 1670 joules. The total power is 28 watts and the intensity at the quartz surface is, then, 8.9 watts per square centimeter.

This value of intensity is only approximate since it assumes that the surface is vibrating uniformly and that the energy dissipated within the crystal did not heat the oil. It also assumes that no heat energy was delivered to the container walls, crystal, and holder. If the electrical energy delivered to the disc is measured at the same time, an approximate value may be had for the conversion efficiency of the crystal. More precise measurements would take into account the effect of the walls and other objects. Once the intensity of a sound field is determined at a given point or for a given geometrical arrangement, it is desirable not only to know its value at any other

time, but also to be able to preset it to some other intensity.   This is
generally done by calibrating the field at the point in question in
terms of the electrical current delivered to a magnetostrictive rod or
the voltage applied to a piezo crystal, whichever source is used.
Once, however, the geometrical conditions or the frequency is changed,
the calibration will change.

C. BIOLOGICAL EFFECTS OF ULTRASONIC VIBRATIONS

Since so many survey articles and résumés (2,3,18,19,24) of the
biological action of ultrasound have appeared in the literature, it will
suffice here merely to mention some of the outstanding experiments
and techniques and to attempt to explain some of the observations.
Unfortunately in some experiments, little if any data were reported
on the sound field used and as a result attempts to determine the
fundamental factor or factors responsible for some biological actions
have met with little success.   First, it will be worth while to reiterate
the properties of a high energy sound field.   Vibration of this sort is
characterized by extremely high pressures and large accelerations of
the particles of liquid in the field.   These in turn produce cavitation
and, if dissolved gas is present, intense local agitation, high local
temperatures, and possibly electrical potentials.   The net result is
an exhibition of lethal and sterilizing effects, strong dispersive power,
degassing processes, thermal and oxidizing effects, and coagulation.

1. Lethal and Sterilizing Action

The first biological effects of high intensity, high frequency sound
waves were observed by Wood and Loomis (11), who reported that
protozoa were paralyzed or killed depending on the length of ex-
posure.   They advanced the concept that all protozoa did not re-
spond alike since the smaller ones could "ride" the waves and not be
greatly affected.   They further felt that rupture of the larger bodies
was probably due to variations in direction and magnitude of the
forces (pressure) applied to different points of the body.   Bacteria
apparently were not affected because the entire organism was sub-
jected to the same force.   Other investigators have supported the ob-
servation that larger protozoa were affected while smaller ones were
not.
    In 1931, Harvey and Loomis (10) made high speed photomicro-
graphs of living cells subjected to ultrasound and found that the de-

struction of *Arbacia* eggs occurred in less than 1/1200 second. Their photographs also showed that the eggs are first drawn out into spindle or tadpole shapes and then are disintegrated. This suggested that rapid fluid movement as the result of submicroscopic cavitation was the main cause of disintegration. A time delay was noted in the destructive action ($\approx 1/300$ sec.), which recently (*22,32*) has been found to be characteristic of cavitation. They also reported that strong eddies first appeared in the cells of the plant *Elodea* with the final result that the protoplasm or chloroplast became separated from the cell wall and coagulated in the other places. Algae were also torn to pieces.

Red blood corpuscles in physiologic saline solution have also been destroyed by ultrasonic radiation. Counts at the end of fifteen second intervals showed that the percentage destroyed decreases as time goes on until no more are affected unless the radiation intensity is increased. Harvey (*32*) found that in all cases laking of blood was associated with the expulsion of gas from a liquid in the form of bubbles (cavitation) and that the type of gas did not matter unless it was extremely soluble in water. That is, air, nitrogen, or hydrogen all worked equally well but carbon dioxide did not. Furthermore, gas bubbles formed before the sound exposure had no effect showing that it was the bubble formation itself during exposure that accounted for the destructive action. He found further peculiarities in the effect of the dissolved gas tension but these seemed to be associated with the phenomenon of cavitation rather than with an intrinsic effect on the organisms. For example, blood was laked under a gas pressure of 100 lb. per square inch as long as the gas was not in equilibrium with the liquid. However, if, at the same pressure, equilibrium were established by thorough shaking of the liquid–gas mixture, a much higher acoustic intensity was required. This supports the general concept of cavitation phenomena as long as the gas–liquid equilibrium tension is considered rather than merely the applied pressure.

In addition to killing many different types of small organisms, larger animals such as fish, frogs, etc., have also been reported killed or lamed in the sound field. This is to be expected since the sound field would be present throughout their bodies and the ensuing destruction of some critical cell or group of cells would naturally produce death or paralysis. As regards the death of larger animals, Harvey (*32*) has found in the case of fish that the major cause is the destruction of the gill tissue and the blood corpuscles.

Concerning the destructive action of ultrasound, a general impression held by a few is that pressure differences still play an important role.   While cavitation cannot exist without large pressure differences, it remains that if the pressure differences are largely responsible for the destructive action there should be a great dependence on frequency—that is, the sound could be "tuned" to the dimensions of the undesired organism.   No data on such an effect have yet been reported.   On the other hand, if cavitation is principally responsible for the destructive action, the elimination of this phenomenon should greatly diminish the virulent action.   Observations of this sort by Schmitt and Uhlemeyer (12) were made on protozoa by maintaining the same acoustic intensity and increasing the external pressure until cavitation and bubble formation ceased.   The destruction of the protozoa was considerably lessened.   Johnson (13) and Harvey and Loomis (10) also made the same observation.   Stanley (14) used a different technique when investigating the inactivation of tobacco mosaic virus by ultrasound.   He found no inactivation in the absence of the dissolved gases.   In this case, cavitation was still present but due to the absence of bubble formation the local agitation and temperatures were not.   It is important to remember that cavitation also has a seeming frequency effect (very slight, however) since the cavities are formed in the trough of the wave.   If cavitation is responsible, this may explain in part the observations of Wood and Loomis and other investigators concerning the ability of smaller protozoa to "ride the waves."   Harvey (32) has made a rather critical investigation of these biological effects and has arrived at the conclusion that most of the lethal and sterilizing action of ultrasound on small organisms is due to rapid fluid movements as a result of submicroscopic cavitation in the presence of dissolved gases.   He has also shown, particularly in the case of plant cells, that cavitation and the expansion of cavitated gases within the cells should have little effect.

Further results of his investigations were that none of the phenomena investigated could be connected with local heating as long as the average temperature of the medium did not rise.   This was prevented by cooling the oil above the crystal and by electrically shielding the medium so that radiofrequency heating (due to the electrical signal applied to the crystal) did not take place.   He further found that the standing sound wave pattern was quite important and determined to a great extent the observed phenomena—particularly the movements and distribution of the particles in the medium.

The general conclusion arrived at from an analysis of all the data to date substantiates Harvey's observations, namely, that lethal and sterilizing effects on small organisms occur only when there is cavitation in the presence of dissolved gases and that cavitation does not occur within the cells but rather causes external fluid movements etc., which in turn tear the cells apart.

Bacteria and other extremely small pathogenic organisms are affected in many different ways. Williams and Gaines (15) reported that *Escherichia coli* were killed and that the lethal effect was probably due to violent action set up within the cell. They used a low frequency source of 8800 cycles per second, which has a wavelength in water of about 17 cm. Since the wavelength in this case is very large compared to the cell size, the lethal action of the sound was probably due to some secondary effect of cavitation—such as local fluid movements, high local temperatures, local potentials, or the production of hydrogen peroxide. The latter effect has been shown to exist by the oxidation of potassium iodide when irradiated in a solution containing dissolved air. Use of another dissolved gas would eliminate the latter factor but no data on such an effect with bacteria have yet been reported. As far as the effect of fluid movements are concerned, it is well known that a rapidly revolving (2000–3000 r.p.m.) spindle with vanes will break up a suspension of cells and sand. Bacteria are destroyed this way in about five hours and while it is possible that exposure to a cavitating ultrasonic field is a more violent exhibition of this phenomenon, it does not explain why some bacteria remain completely refractory to ultrasonic treatment.

Some investigators have reported increased agglutination and diminution of virulence when certain bacteria are irradiated with ultrasound, while others have reported an opposite effect—for example, that colon bacilli could not be influenced, no matter how long the exposure or how high the intensity. Yeast cells have been found to lose their power of reproduction and luminous bacteria their luminosity. All in all, the effect on bacteria, because of their small size, seems to depend on the reaction of the particular organism to the secondary effects of cavitation. While this represents a promising field in biology, ultrasound does not seem to have the sterilizing action popularly ascribed to it.

Stumpf, Green, and Smith (31) have investigated ultrasonic disintegration as a method of extracting bacterial enzymes. They used a quartz crystal generator (29) with a maximum power of 700 watts

at 600 kilocycles. The crystal was mounted with an air backing in an oil bath and the bacterial suspensions were introduced into the sound field by means of a thin-bottomed flask. After a given exposure time, the suspensions were centrifuged and analyzed. They found that the number of bacteria disintegrated increased with exposure time, leveling off somewhat at 56% disintegrated for ten minute exposure and 63% for fifteen minute exposure. A ten minute exposure was accepted as a reasonably standard interval for subsequent measurements and extractions. They further found that the viscosity of the suspension and the shape of the container greatly influenced the results. In particular, thick suspensions and pastes showed very little disintegration even though they became heated. In this process, heat has no disruptive effects.

While they made no observation of the actual sound pressure in the suspensions or cavitation conditions, a plausible explanation is that the sound pressures involved were large enough to produce cavitation in the thin suspensions and not in the much more viscous ones. As mentioned in the section on cavitation, heavy viscous liquids require two to four times the sound pressures to produce cavitation due to their larger cohesive pressures. They also found that the degree of disintegration was independent of frequency and dependent only on the intensity—a characteristic of cavitation. If the power were raised beyond a certain point, denaturation and inactivation of the labile proteins occurred rapidly so that ultrasonic disintegration (production of cell-free enzymes) took place only in a very narrow range of power. Investigation of various types of bacteria showed that some were easily disintegrated while others were completely refractory.

Recently some patents have been granted (*33*) on frequency modulating a sound field by moving a reflector back and forth in front of the sound radiator. Since the moving reflector constitutes a second source of the primary ultrasonic radiation, it follows that the wavelength of the sound in the medium will be changed alternately from a certain maximum value to a certain minimum value due to the Doppler shift in frequency by a moving source. While this was proposed as a means of irradiating organisms with various frequencies simultaneously in case there were critical lethal frequencies, the latter have not been shown to exist. However, this technique has been used previously (*27*, pp. 46 and 112) to eliminate standing waves in a volume filled with sound energy and might prove of value in biological

and chemical investigations in which the nodes of standing waves would produce no action. Stumpf, Green, and Smith (*31*) found that the shape of the container affected the rate of bacterial disintegration, a direct exhibition of this phenomenon, and it is felt that this may also explain to a great extent the earlier observations that small protozoa could "ride the waves" while the larger ones could not, the "waves" being the nodes and antinodes of a stationary sound field pattern.

Instead of employing mechanically moving reflectors, which might be cumbersome to use in any given investigation, it is possible to frequency modulate the sound source directly by standard electronic circuits (*34*). These should prove much more adaptable particularly at the higher frequencies. Offhand, it appears that the possibility of frequency modulating the sound source should provide a very powerful tool in a more complete investigation of biological and chemical action of ultrasonics in small containers, where the effects of the boundaries become important. The influence of the boundaries on the resultant sound field pattern and the consequent effects cannot be emphasized too strongly.

2. Thermal Effects

When dealing with investigations of this sort, it is important not to overlook the heating effects of the sound. There are in general two main effects: high local temperatures due to cavitation in the presence of dissolved gases and a general increase in temperature due to the absorption of the sound energy by the medium.

Porter and Young (*16*) have reported the conversion of an acid azide into an isocyanate, which is a known case of readjustment of atoms within a molecule by vibrational energy or heat. A thermometer immersed in the solution showed a change of 0.3°C. during the run while a temperature of 90° was necessary to decompose the azide at the rate measured. It is a plausible assumption that the necessary temperatures were local phenomena generated by cavitation. These local heats would not appreciably raise the temperature of the over-all liquid but might increase tremendously the reaction rate at the points where the local temperatures were high. Marinesco (*17*) further proved this by detonating such labile materials as NI_3 with ultrasound. No explosions occurred when the liquids were degassed.

Many organic materials show a large general rise in temperature due primarily to the absorption of the radiation and not to the effect

of cavitation. Dognon and Biancani (*18*) have investigated this rather thoroughly and have arrived at a general rule for this effect in fatty bodies and liquid or coagulated proteins. Tables III, IV, and V from Pohlmann (*30*) and Dognon and Biancani (*18*) show the results of such investigations. Table III shows that the temperature rise is less in a gelatin solution than in water even though gelatin is expected to have a higher absorption. This is probably due to the

TABLE III

Rise in Temperature in 2 Ml. of Various Liquids for a Ten Second Ultrasonic Exposure

Liquid	ΔT, °C.
Water	2
Alcohol	3.5
Glycerol	10
Liquid paraffin	10
Gelatin solution	1
Gelatin gel	1
Stearic acid	36
Wax	44

TABLE IV

Temperature Rise in Various Materials after a Thirty Second Ultrasonic Exposure

Material	ΔT, °C.
Agar gel	0
Brain	9.5
Coagulated egg albumin	2
Egg albumin	2
Egg yolk	11
Fat	25
Liver	9

TABLE V

Absorption Coefficients and Half-Value Thicknesses for Various Tissues

Material	Frequency, kilocycles	Absorption, $(\alpha/f^2) \times 10^{15}$ cm.$^{-1}$ sec.^{2}	Half-value thickness, cm.
Layer of fat	800	80	6.8
Layer of muscle	800	150	3.6
Fat plus muscle	800	110	4.9
Fat plus muscle	2400	41	1.5

absence of bubble formation in the gelatin solution, which would re-
sult in smaller absorption even with greater viscosity. One other
interesting point is the change with frequency in α/f^2 for fat tissue
plus muscle (Table V). This does not occur for most pure substances.

Other investigators have found that the marrow of the bone may
be heated without affecting the bone and have proposed this as being
useful in medical diathermy. Pohlmann (30) has designed a hand
applicator (see Fig. 8) in which the vibrating crystal is held directly
against the skin for sciatic and plexus neuralgia. He further recom-
mends the use of a contact substance such as paraffin oil or linament
between the massager and the skin. Excellent results have been re-
ported for such treatment.

Regarding production of heat by absorption, Lynn, Zwemer, and
Chick (21) reported interesting results with a focused quartz crystal
at a frequency of 835 kilocycles. The crystal was ground to provide
a definite focal point for the sound waves emitted by the surface (see
Sect. B2 and Fig 1E) and the sound passed out through a Cellophane
diaphragm so that direct contact could be made with the tissue. In-
tense heating at the focal point was shown by exposing blocks of
paraffin and liver to the sound beam. The paraffin melted and the
liver "cooked" white at the focal point. Examination with a micro-
scope showed destruction of the cells at the focal point and in the path
of the beam.

In experiments with living animals, production of lesions is made
difficult by circulating blood. For this reason, the brain of a dog was
chosen since any destructive effects would become evident by a dis-
turbed motor activity as soon as the effects of the anesthetic dis-
appeared. Despite the fact that the focus was disturbed by bone re-
fraction and scattering, positive results were obtained. With trans-
cranial application of the focused sound beam to the proper areas of
the cerebral cortex, disturbances in muscular coordination, some
paralysis, and, in one case, blindness were observed. These were
associated with the corresponding brain lesions. Necrosis of the
scalp also occurred where the apparatus was applied.

3. Chemical Effects of Ultrasonics and Secondary Results of Cavitation

The acceleration of chemical reactions by ultrasonics has been
shown to be due to cavitation in the presence of dissolved gases.
The majority of these reactions are oxidations such as the conversion

of iodine ions into iodine molecules but depolymerization, inversion of sugar, and effects on hydrolysis have also been reported. The most plausible explanation to date is that most of the effects are due to high local temperatures generated by dissolved gases filling the small cavities. This, as was mentioned in the preceding section, was substantiated by the independent experiments of Porter and Marinesco.

The above phenomena, however, can explain neither luminescence nor the formation of hydrogen peroxide from oxygen and water, which appears to be the basis of the acceleration of some chemical reactions. This has been brought out by the fact that many reactions are accelerated by the cavitation of certain dissolved gases (air and oxygen) while they are unaffected by others. In order to overcome this difficulty, Richards (19) has proposed a balloelectrical theory. This theory is based on the concept that just as spraying a liquid into a gas produces charged drops, the spraying of holes in a liquid (cavitation) must produce charged bubbles. The high local potentials might then be the cause of the various chemical reactions. Measurements on cavitating fluids showed large variations in potential at the points of cavitation and none where the liquid was relatively undisturbed. This theory also accounts for the fogging of photographic plates when exposed to ultrasonic sound. Needless to add, both the high local temperatures and potentials may be present at the same time in any cavitating fluid so that an analysis of any given chemical or biological effect may involve one or both phenomena.

Other physicochemical effects of ultrasonics are the alteration of sensitive metastable systems such as the yellow to red change in mercuric oxide at room temperature and the detonation of explosives by vibrating glass surfaces in air.

4. Emulsification and Dispersion

One of the first things demonstrated by intense ultrasound was the transformation of two immiscible liquids into a very stable emulsion (11). Water and oil, and water and mercury were first tried and since then a variety of mixtures has been used. Richards (19) has shown that emulsification is very strong at the boundary surfaces between the liquid and vibrator and also between the liquid and walls of the vessel. Other investigators have shown that the emulsions produced are different depending on whether or not a gas is present

as a third phase. Emulsions containing gas are not only much easier to produce, but also much more stable.

In a similar manner, low melting metals and alloys have been dispersed in liquids. The particle size averages about 0.5 μ. Another aspect of this effect of ultrasonics is the decrease in grain size of metals with treatment of the crystallizing melt. This is to be expected with all crystallizing substances since the ultrasonic agitation breaks off small particles of the growing crystals, which in turn act as nuclei. This dispersive effect is also being used in preparing certain pharmaceutical compounds. For example, Chambers and associates (20) have been able to disperse sulfathiazole into such fine particles that, when suspended in a water or a salt solution, a cream is produced that can be injected through fine hypodermic needles. The emulsions produced by ultrasonics are often much more stable than those produced by colloid mills or similar devices. There is a series of technical products appearing commercially, ranging from cosmetics to photographic plates, that can be produced with higher quality by ultrasonic emulsification than any other way.

5. Coagulation Effects

While ultrasonics have a strong dispersive effect on emulsions and liquid sols, they have an opposite effect on aerosols, that is, they exhibit a strong coagulating action on liquid or solid particles suspended in a gas. This coagulating action has received impetus lately in attempts to precipitate fog, smoke, and industrial waste. Since a new medium is involved, large generators designed primarily for air operation have been developed. Allen and Rudnick (25) have reported an air siren capable of delivering a power exceeding 2 kilowatts over a frequency range of 3 to 19 kilocycles. With the sound energy from such a siren, they have been able to ignite cotton wads in six seconds and steel wool in sixty seconds. In the case of the cotton, only the first inch was burned, showing that the acoustic energy was almost totally absorbed in that thickness. It was also possible to heat a beaker of water from room temperature to boiling in about seven minutes. Paper and cloth were also shredded in this sound field. When the siren was adjusted to produce a vertical standing wave field, small objects such as marbles ($\frac{3}{4}$ inch in diameter) and small coins floated at the velocity antinode surfaces.

The sound field also produced a burning sensation in the hand when

held in the field with the fingers closed.   When the fingers were held
open, the sensation was relieved.   A temperature increase at the hand
of 45°C. was measured with a thermocouple while glass and rubber
showed only a 1° rise.   Their concept is that the heating is due to the
damping (absorption) of vibrations, not necessarily of the same fre-
quency as the sound, that are set up in the fingers by the intense sound
field in the region between them.

A study of the effects of these amounts of power on airborne bac-
terial and larger organisms should prove both interesting and fruitful.

6. Natural Sources of Ultrasonic Sound

The production and use of ultrasonic sound is by no means con-
fined to the laboratory.   Fish, shrimp, grasshoppers, crickets, birds,
and bats all produce acoustic energy in the ultrasonic frequencies.
In the case of the grasshopper, for example, energy has been de-
tected as high as 40 kilocycles.   As far as fish and shrimp are con-
cerned, while some energy has been detected in the higher frequencies,
most of the energy is concentrated in the audible range at about 2.5
kilocycles.   Some birds, the canary is a notable example, actually sing
at these higher frequencies (on the order of 20 kilocycles), while other
animals such as cats and dogs can only detect and not produce them.
While there is some question as to what benefit most animals derive
from ultrasonic frequencies, the bat actually depends upon them for
navigation in flight in a manner similar to ultrasonic (sonar) and
radar echo ranging devices.

Galambos and Griffin (26) have shown that the bat actually pro-
duces three types of sound.   One, a shrill cry of anger, is usually at a
frequency of about 7 kilocycles.   The second, a series of clicks, is
associated with the production of the third type of sound—a series
of pulses (sound wave trains) at a frequency of about 70 kilocycles.
Each pulse lasts about 0.01 second.   If the bat is at rest, only about
five to ten clicks per second are observed.   When in flight, this in-
creases to twenty to thirty per second and, in the vicinity of objects,
to fifty per second.   At these frequencies (70 kilocycles), the sound is
not only beamed from the source (seemingly the larynx), but also is
reflected quite well from nearby objects.   The time between sending
and receiving a pulse obviously gives the bat a measure of the distance
to the reflector, while the number of pulses per second increases his
accuracy and speed of detection at close range.

Some animals, on the other hand, are affected by ultrasonic frequencies even though they do not produce them. Pigeons, for example, are repelled by 20 kilocycle sound so effectively that this is being considered as a means of driving them away from public buildings where they have proved themselves a nuisance. The same effect has been observed in cats. On the other hand, it appears that simple marine life like the jellyfish is attracted by intense underwater sound. Man is also not immune. Various investigators (15) have reported such effects as headaches, loss of balance, loss of the ability to perform various mathematical computations, and a lowering of the visual light threshold after exposure to ultrasonic sound (28).

7. Miscellaneous Applications of Ultrasonics

As far as the biological applications are concerned—the physical applications are another field in themselves (2,3,19)—much has been proposed and accomplished along the lines of food preparation and sterilization. In some cases, it has actually been possible to sterilize inside the can in which the product has been sealed. Dairies use a simple vibrating diaphragm that homogenizes milk and kills most of the bacteria in it. The net result is a milk that forms smaller curds in the stomach and is recommended for babies. Further, the bacterial count of milk has been reduced by such treatment to eight per cubic centimeter while a count of 30,000 indicates a high standard for pasteurized milk treated in the conventional thermal manner. Bacterial disintegration by ultrasonic sound has also been used in the preparation of endotoxins and enzymes. Further use has been made of the emulsifying, coagulating, and dispersive properties of sound in the preparation of pharmaceutical compounds, mayonnaise, peanut butter, paints, chemicals, and even in the preparation of cosmetics. It has also been used in the aging of liquor. An ultrasonic process has been suggested for the aging and curing of meat.

As far as plant life is concerned, ultrasonic sound has been used to accelerate and stimulate germination. Some investigators have reported that a potato crop flowered a week earlier than control plants, and with an increase in yield. Peas sprouted earlier also.

One of the more interesting recent developments is a "seeing eye" device for the blind. This instrument is essentially an adaptation of underwater echo ranging equipment to airborne sound. Because of the high "beaming" or directional characteristic of high frequency

sound, it is possible to send out a short train of sound waves and to receive a reflection back from nearby objects that are in the same direction as the projected sound. The time interval between repeated pulses determines the maximum range of the device while the time interval between pulse and echo determines the distance to the object. A crystal transducer in this case produces adequate intensity and sensitivity. The major problem has been to translate the received pulses into some recognizable signal. This has been done in one system by frequency modulating the sound source in such a manner that the addition of the received signal to that being transmitted at that instant produces an audible tone proportional to the distance. Objects and discontinuities such as curbs have been detected successfully up to about fifty feet.

References

1. Wood, A. B., *A Textbook of Sound*. Macmillan, New York, 1930.
1a. Beranek, L. L., *Acoustic Measurements*. Wiley, New York, 1949.
2. Bergman, L., and H. S. Hatfield, *Ultrasonics and Their Scientific and Technical Applications*. Wiley, New York, 1938.
3a. Hiedemann, E., *Ultraschallforschung*. DeGruyter, Berlin, 1939.
3b. Carlin, B., *Ultrasonics*. McGraw-Hill, New York, 1949.
3c. Fry, W. J., J. M. Taylor, and B. W. Henvis, *Design of Crystal Vibrating Systems for Projectors and Other Applications*. Dover, New York, 1948.
3d. Camp, L., *J. Acoust. Soc. Am.*, **20**, 289, 611, 616 (1948) (magnetostrictive transducer designs).
3e. Dranetz, A. I., G. N. Howatt, and J. W. Crownover, *Tele-Tech*, **1949**, May, p. 28, and June, p. 36 (barium titanate as a piezoelectric crystal element).
4. Cady, W. G., *Piezoelectricity*. McGraw-Hill, New York, 1946.
5. Pierce, G. W., *Proc. Am. Acad. Arts Sci.*, **63**, 1 (1928); *Proc. Inst. Radio Engrs. and Waves and Electrons*, **17**, 42 (1929).
6. Salisbury, W. W., and C. W. Porter, *Rev. Sci. Instruments*, **10**, 142 (1939); **10**, 269 (1939).
7. Greutzmacher, J., *Z. Physik*, **96**, 342 (1935).
8. Tumanski, S. S., *J. Tech. Phys. U. S. S. R.*, **7**, 2019 (1937).
9. Smith, A. W., and D. K. Weimer, *Rev. Sci. Instruments*, **18**, 188 (1947).
10. Harvey, E. N., and A. L. Loomis, *J. Gen. Physiol.*, **15**, 147 (1931).
11. Wood, R. W., and A. L. Loomis, *Phys. Rev.*, **29**, 373 (1927).
12. Schmitt, F. O., and B. Uhlemeyer, *Proc. Soc. Exptl. Biol. Med.*, **27**, 626 (1930).

13. Johnson, C. H., *J. Physiol.*, **67**, 356 (1929).

14. Stanley, W. M., *Science*, **80**, 339 (1934).

15. Williams, O. B., and W. Gaines, *J. Infectious Diseases*, **47**, 485 (1930).

16. Porter, C. W., and L. Young, *Calif. Eng.*, April, 1938; *J. Am. Chem. Soc.*, **60**, 1497 (1938).

17. Marinesco, N., *Compt. rend.*, **201**, 1187 (1935).

18. Dognon, A., and H. Biancani, *Radiologica Berlin*, **3**, 40 (1938); *Ultrasons et Biologie*, Gauthier-Villars, Paris, 1937.

19. Richards, W. T., *J. Am. Chem. Soc.*, **51**, 1724 (1929); *Rev. Modern Phys.*, **11**, 36 (1939).

20. Chambers, L. A., T. N. Harris, F. Schumann, and L. K. Ferguson, *Science*, **95**, p. 11 in suppl. (June 5, 1942).

21. Lynn, J. G., R. L. Zwemer, and A. J. Chick, *Science*, **96**, 119 (1942), *J. Gen. Physiol.*, **24**, 179 (1942).

22. Briggs, H. B., J. B. Johnson, and W. P. Mason, *J. Acoust. Soc. Am.*, **19**, 664 (1947).

23. Willard, G. W., *J. Acoust. Soc. Am.*, **12**, 438 (1941); *Bell Labs. Record*, **25**, 194 (1947).

24. Gregg, E. C., in O. Glasser, *Medical Physics*. Year Book Publishers, Chicago, 1944, p. 1591.

25. Allen, C. H., and I. Rudnick, *J. Acoust. Soc. Am.*, **19**, 857 (1947).

26. Galambos, R., and D. Griffin, *J. Exptl. Zool.*, **89**, 475 (1942); **86**, 481 (1941).

27. *Sonar Calibration Methods*, Summary Technical Report of Division 6, National Defense Research Committee, Volume 10.

28. Kektscheev, K. C., and P. Ostrovski, *Compt. rend. acad. sci. U. S. S. R.*, **31**, 370 (1941).

29. Smith, F. W., and P. K. Stumpf, *Electronics*, **19**, 116 (1946).

30. Pohlmann, R., *Physik. Z.*, **40**, 159 (1939); *Forschungen u. Fortschr.*, **15**, 187 (1939); *Deut. med. Wochschr.*, **65**, 251 (1939).

31. Stumpf, P. K., D. E. Green, and F. W. Smith, Jr., *J. Bact.*, **51**, 487 (1946).

32. Harvey, E. N., *J. Cellular Comp. Physiol.*, **24**, 1 (1944); *J. Am. Chem. Soc.*, **67**, 156 (1945); *Biol. Bull.*, **59**, 306 (1930); *J. Bact.*, **27**, 373 (1929).

33. U. S. Pat. 2,424,357 (liquids) and 2,424,375 (air); see also *Science News Letter*, **52**, 89 (Aug. 9, 1947).

34. Terman, F. E., *Radio Engineers' Handbook*. McGraw-Hill, New York, 1943.

35. Sollner, K., "Sonic and Ultrasonic Waves in Colloid Chemistry," in J. Alexander, *Colloid Chemistry*, Vol. V. Reinhold, New York, 1944, pp. 337–370.

36. Novotny, H., *Werkstoffzerstörung durch Kavitation*. VDI Verlag, Berlin, 1942 (reprint by Edwards, Ann Arbor, Michigan, 1946).

37. Pease, D. C., and L. R. Blinks, *J. Phys. Colloid Chem.*, **51,** 556 (1947). (cavitation at solid surfaces).

38. Oster, G., *J. Gen. Physiol.*, **31,** 89 (1947) (ultrasonic treatment of tobacco mosaic virus).

COMMERCIAL SOURCES OF ULTRASONIC EQUIPMENT

39. Eimer and Amend, New York; Fisher Scientific Company, Pittsburgh; Televiso Products Company, Chicago, Ill.; Piezo Products Company, Framingham, Mass.

40. Raytheon Manufacturing Company, 178 Atlantic Ave., Boston 9, Mass.

CHAPTER XI

WHEN TO USE SPECIAL MICROSCOPES

Oscar W. Richards, *American Optical Co.*

A. PROPERTIES OF MATERIALS ESSENTIAL
FOR MICROSCOPY

By means of the microscope, the identification of a material or an impurity can be accomplished without damage to the material itself and by the use of an extremely minute sample. Since the nature and structure of materials can be discovered so advantageously with it, the microscope is considered a necessary experimental tool in almost every laboratory. Actually, microscopy comprises many kinds of instruments and techniques. Since some of these require considerable equipment, skill, and training they will be discussed at greater length, but first some general principles will be established.

The simple microscope, or magnifying glass, may be any spherical lens thicker at the center than at the edge. Magnifiers are most useful when the desired detail may be seen with 15 diameters ($15\times$) magnification or less. For greater magnification the compound microscope is used. Two lens systems are used in the compound microscope, one of which (ocular or eyepiece) magnifies the image produced by the other (objective) and the total magnification is the product of the two magnifications referred to a plane 10 inches from the eyepoint. While the image might be magnified further by more and more lens systems, actually some light is lost and the image deteriorated due to residual defects at each lens system. Hence, no practical gain is derived from a combination of more than two lens systems.

1. Image Formation—Possibilities and Limitations

The information obtained from the use of a microscope must come from the examination of the enlarged image formed by it of the specimen. Therefore it is advantageous to give some consideration to image formation and to the limitations of microscopy in general before evaluating the different instruments and techniques. The limitations of the microscope derive from: the available probing radiation, the nature of the materials used in its construction, the difficulties in preparing specimens for examination, the sensitivity of the observer's eye or other recording procedure, and the difficulty of interpreting a greatly enlarged image of a specimen.

Electromagnetic radiation is used ordinarily for microscopy, but electrons behave similarly and may be used (cf. Chapter XII) and even a proton microscope has been described (15). The fineness of

detail revealed by brightfield methods depends in part on the wavelength of the radiation used. Visual microscopy is limited to light of about 380 to 740 mμ because of the sensitivity of the human eye. Longer and shorter wavelength radiation can be used with photographic and other recording mechanisms.

The image formed by a lens of a small point of light is a central disc surrounded by a series of dark and bright rings. A good lens concentrates most of the light into the central disc as shown by the plot of the Airy disc image in Figure 1B. As two such points in the object approach and gradually merge, their images overlap and separate
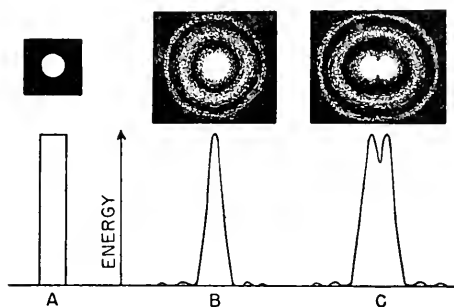
Fig. 1. Light distributions in the object (A), the image (B), and for the resolution of two objects (C).

identity is lost when there is no dip, as in Figure 1C, between the fusing energy curves. The limit of resolution of the microscope is thus determined in terms of distance. The image of the specimen is the sum of the image discs from all the points of light from the specimen. Enough energy must be present in each disc to stimulate the eye or recording mechanism, otherwise it will not be perceived. This is important when the energy distribution is quite different in adjacent Airy discs in the image. Unless the energies in both Airy discs are adequate, methods for increasing resolution (16) should be used with care to prevent incorrect observations.

Resolution, or the minimum separation of two small objects that still permits them to be observed as two, depends on the wavelength (λ) of the radiation used and the angle of the cone of light from the specimen accepted by the front lens of the objective. The light-gathering power of an objective is expressed as *numerical aperture* (N.A.) which is the sine of the half-angle (U), of the above cone, multiplied

by the lowest refractive index (n) of the medium with which it is used, $e.g.$, air, water, or oil of the same index as the front lens of the objective (Fig. 2). The resolving power of an objective is $\lambda/(2n \sin U)$, or $\lambda/(2 \text{ N.A.})$, when its aperture is filled with light by a condenser. The resolving power for white light is usually calculated with $\lambda = 555$ mμ, the region to which the eye is most sensitive. This formula gives the Abbe limit of resolution and modern research indicates that resolution may be better than this value, perhaps 80% of it. Note that this value is for the separation of two points.

With an objective of N.A. $= 0.85$, the preceding formula gives a limit of resolution of 0.47 μ for infrared radiation of 800 mμ, of 0.33 μ for light at the maximum sensitivity of the eye at 555 mμ, and 0.2
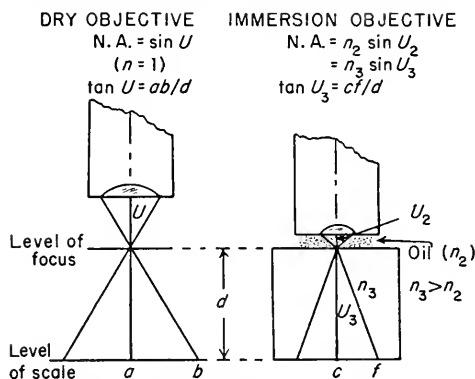
Fig. 2. Numerical aperture and its measurement.

μ for ultraviolet of 257 mμ wavelength. With N.A. $= 1.4$ the limit for 555 mμ light is 0.2 μ. The greatest numerical aperture available is 1.6, but its use requires not only slides, but also mounting and immersion media of correspondingly higher index than ordinary crown glass; it has not proved generally useful. Such a high aperture lens also has a shorter working distance. When the approximate size of the detail to be examined is known, the above limiting values can be an aid in selecting the objective to be used and in the decision whether to use a light or an electron microscope.

While resolution is important and necessary, it is not sufficient for observation. The image must be bright enough to stimulate the recording medium, adequate contrast must exist, and magnification must be great enough to match the resolving power of the observer. For the eye a magnification of about one thousand times the numeri-

cal aperture of the objective used is desirable. Much greater magnification fails to reveal new detail and is referred to as empty magnification. Since the brightness of the image varies directly as the square of the numerical aperture but inversely as the square of the magnification, i.e., as $(N.A.)^2/M^2$, too great magnification may give so little light as to make the image scarcely visible. The same rule is useful for photomicrography, although the necessary magnification will depend on the resolution of the emulsion and the viewing distance at which the print will be examined (7). Some excess or empty magnification may be helpful in measurement.

Monochromatic light passing through the edge of a simple magnifying lens is focused closer to the lens than that passing through the central part of the lens. This variation from a single focal point is called spherical aberration. Chromatic aberration arises when long wavelength red light comes to a focus farther from the lens than shorter wavelength blue light. To correct these and other aberrations (coma, astigmatism, curvature of field, and distortion) positive and negative lenses of various kinds of glass or crystalline materials are combined. Since available optical materials do not permit perfect simultaneous correction of all aberrations, each manufacturer must list many types of lenses. Each type is designed to be the best possible compromise for some specific application. To be capable, a microscopist must know which lenses are best for a given problem.

When a microscope is focused on a reasonably transparent specimen, the object is not all seen at once, but only section by section because of the limited depth of field of the objective. The structure of the specimen in depth is comprehended by slowly focusing up and down throughout its depth. Lenses of lower numerical aperture have greater depth of field. Lenses of great numerical aperture have very little depth of field. The latter are useful for *optical sectioning* (17).

Few specimens occur naturally in suitable form for examination with the compound microscope. They must be thin enough to transmit light or reflect enough light to be seen, and of suitable shape for manipulation. Many methods have been developed for the preparation of the specimen, e.g., sectioning, staining, isolating, and orienting the specimen. Some specimens require high skill and special equipment. Many books give detailed instructions and should be consulted (6,9). Unless the specimen has been properly prepared, its examination with a microscope is likely to be useless.

2. Nature of Specimen Determines Choice of Microscope and Method

Knowledge of how the specimen may affect the probing radiation used for the formation of the image leads to the choice of the appropriate microscope and the best method for the most efficient examination of the specimen. Image formation may depend on one or a combination of the following properties of the specimen: (*a*) Transmission: uniform or selective by wavelength. (*b*) Reflection. (*c*) Optical path effects: refraction, retardation. (*d*) Diffraction. (*e*) Scattering. (*f*) Polarizing, rotation of the plane of polarized light; pleochroism. (*g*) Fluorescence and phosphorescence.

Usually a specimen is partially transparent, or may be cut or ground thin enough to transmit enough radiation for its examination. Few specimens are uniformly transparent to all wavelengths. Most specimens absorb some wavelengths of radiation more completely than others and when this selective absorption is in the region to which the eye is sensitive (380 to 740 mμ) color results. Colored and gray images are visible and such a specimen would be examined with the *brightfield* microscope (Sect. B). When the selective absorption occurs for shorter wavelength radiation the *ultraviolet* microscope should be used and when of longer wavelength the *infrared* microscope is the instrument of choice (Sect. C). When the absorption is weak, *phase* microscopy will be helpful (Sect. I). Opaque and nearly opaque specimens can be examined by reflected light using *vertical* and *epi* methods of microscope illumination (Sect. D). The *bi-objective* microscope (Sect. D2) reveals three dimensional form at lower magnifications and for the higher magnifications the *brightfield* compound microscope is used.

Quite transparent objects usually contain regions of different refractive index and size, which alter the direction and speed of light passing through them. The product of the thickness (t) of the region times its refractive index (n) is the *optical path*, which is sometimes expressed in wavelengths (nt/λ). When these path differences are small, *phase* microscopy (Sect. I) will reveal otherwise invisible detail, but when the optical path differences are large the *darkfield* methods (Sect. E) will be more useful to the microscopist. Surface detail of some specimens and nonuniformities of optical path in transparent specimens may be observed with the interference microscope (Sect. H). Diffraction contributes to most image formation and a

pure diffraction image is formed when the specimen is smaller than the limit of resolution of the objective used.   If the illumination is sufficiently intense the bright central part of the diffraction disc (Fig. 1) may be seen with *darkfield* or *phase* microscopes (Sect. E1, Sect. I) and some knowledge gained of the average size and distribution of the particles, even though the size and shape of the individual particle may not be measurable.

When the specimen orients the vibration direction of the light passing through it, or rotates the plane of polarized light passing through it, the *polarizing* microscope should be used (Sect. F).

Some specimens will absorb radiation of one wavelength and re-emit the energy at another, usually longer, wavelength.   This is called fluorescence.   Also, objects may be treated with a fluorescent chemical, a procedure analogous to staining, and examined by *fluorescence* microscopy (Sect. G).   Phosphorescing specimens continue to emit light for a short time after the exciting energy is stopped and a microscope has been devised for work with such specimens (Sect. G).

Rarely is the image seen in the microscope due to a single one of the above causes.   Even so, when one image type predominates, the kind of microscope indicated becomes the instrument of choice. When the specimen gives several types of image, it may be necessary to use several or all of the microscopes and methods to fully comprehend the specimen.   In general these methods are complementary and the objective of the investigation will assist in the choice of instrument.   Since it is impractical to cross-index the interrelations of all these methods in this section, it is suggested that the investigator read the entire chapter for a broad over-all concept of the possibilities of microscopy before selecting the procedure to be used in a given investigation.

B. COLORED SPECIMENS—BRIGHTFIELD MICROSCOPY

Specimens selectively absorbing or transmitting wavelengths of light to which the eye is sensitive have color.   The image from such a specimen is likewise colored and the color aids in recognizing the fine structural detail.   Selective staining of specimens with dyes may differentiate and make possible the observation of detail that would otherwise not be visible.   The ordinary brightfield microscope is used for the examination of colored objects.

Achromatic objectives are corrected for chromatic aberrations

(*cf.* Sect. A1) at two wavelengths or colors and spherical aberration at
one wavelength. They are satisfactory for ordinary microscopy.
The further correction of aberrations in the objective requires the use
of fluorite as well as glass. Apochromatic objectives are corrected
chromatically for three colors and spherically for two colors. The
fluorite, or semiapochromatic, objectives are of intermediate correc-
tion. Apochromatic objectives give better images and are desirable
for the examination, and especially for the photomicrography, of
colored specimens. Apochromatic objectives require achromatic–
aplanatic condensers and compensating oculars for best image forma-
tion.

Limiting the light with a suitable filter to the yellow-green will
give a better image with achromatic objectives (when this color
limitation is not objectionable), because that is the region for which
the objectives are corrected and also the region of greatest sensitivity
of the eye.

The contrast in the image of colored specimens may be increased
or decreased by varying the color of the light used to illuminate the
microscope with filters. Specimens may be stained with several dif-
ferent colors for greater differentiation of structural detail. The
methods are well known and described in reference books (*2,4,6*).
Increased color contrast may also be obtained by optical staining and
with phase microscopy (Sect. E3, Sect. 1). Color is also visible
when the preparation is observed with the darkfield microscope (Sect.
E1) when contrast of the color against a dark background is desirable.

Color filters are available in glass and gelatin to meet most needs
(*61,62*). Tungsten light contains a greater proportion of the longer
wavelengths than sunlight and may be changed to daylight quality by
filtering it through a blue glass. The nature and thickness of the
filter must be balanced to the color temperature of the lamp and the
lamp maintained at the design voltage. For microscopy, light corre-
sponding to the temperature range 4500 to $5000°K$. gives good color
rendition. For natural color photography, color-compensating filters
are usually necessary and the color should be made correct at the
plane of the film or plate (*7*).

The disadvantages of staining specimens for observation arise
from the time and materials required and the fact that most living
organisms are injured or killed, since very few stains are nontoxic
enough to be used on living material. Some specimens cannot be
colored with solutions of dyes while others require mordanting or

other chemical pretreatment before staining. Most stains used are relatively permanent, but even so, the preparations should be stored away from light, especially sunlight. Some experience and training are necessary to interpret color contrasts and for the preparation and examination of natural and artificially colored specimens.

C. SPECIMENS THAT ABSORB NONVISIBLE RADIATION

Longer and shorter wavelength radiation that is not visible is useful for microscopy when absorbed by specimens. Special materials are required for the construction of microscopes that are transparent to the invisible radiation used, likewise for the mounting of the specimen. Since the image cannot be seen, it must be observed by photography, photoelectric cells, or other indirect methods. This involves understanding the use of special mechanisms for focusing and recording the image. The microscopist must be familiar also with the use and maintenance of the equipment required for producing the radiation used. For efficient use this may require considerable training and skill. Even so, much useful information has been obtained by the use of infrared and ultraviolet microscopy.

1. Infrared Microscopy

Ordinary microscopes transmit to a considerable extent in the infrared region ($\lambda > 740$ mμ). The resolving power of the microscope decreases as the wavelength of radiation used increases. Most infrared microscopy has been done with radiation of 800 to 1000 mμ. Tungsten lamps produce adequate amounts of this energy. A filter is required to absorb the visible and shorter radiation and the photographic film or plate must be sensitive to the infrared radiation used.

For the near infrared (800 mμ) the microscope may be focused with a trichromatic red filter (Wratten A) and the Wratten 87 filter substituted for it when exposing the picture. Other filters and equivalent filters of other make are listed by Clark (18). Apochromatic objectives usually give better images in the infrared than achromatic objectives. For focusing purposes in the longer wavelength regions, a preliminary calibration curve may be made by exposing a series of pictures at known positions of the fine adjustment. Once the adjustment has been calibrated, the correct focus can be set quite readily. Another method is to note the reading on the fine

adjustment for best focus in the green and in the red and to continue
to turn the fine adjustment in the same direction for a distance equal
to twice the difference between the red and green foci (*18*).   Since
some wooden plate holders and the hard rubber plate holder slides
are transparent to this radiation, metal ones should be used, or others
tested to make certain that they are satisfactory.

Visual observation is possible by means of the 1P25 converter
tube.   The microscope image focused onto the sensitive surface of
the tube releases electrons and the rest of the tube is an electrostatic
microscope which focuses the electrons onto a fluorescent screen.
The fluorescent image can be observed through a magnifier (*18a*).

Red stained specimens and some natural reddish materials reveal
considerable detail with infrared microscopy.   The dark chitin of
many insects and fossil graptolites have been successfully examined
in this manner.   Thicker sections may be used than with shorter
radiation.   Staining the specimen with kryptocyanine or neocyanine
and using photographic film sensitized with these same dyes makes
good use of the absorption bands to reveal any detail selectively
stained.

Infrared microscopy has also been found useful for the examination
of colored coral skeletons, silver-impregnated nerve tissue, kidney
sections, capillaries and injected capillaries in tissue sections, em-
bryos stained with silver nitrate or with carmine, plant cell walls,
woody structures, calcareous algae, and textile fibers.   Too little
infrared microscopy has been done to permit a critical evaluation at
the present time (*18*, p. 275).

2. Ultraviolet Microscopy

Proteins, nucleic acids, and other interesting constituents of bio-
logical materials have absorption bands in the ultraviolet region,
which permit the microscopic identification and quantitative meas-
urement of these substances within single cells.   In addition the in-
creased resolution resulting from the shorter wavelength has made
the method useful for examination of those materials that were just
below the limit of resolution with the ordinary microscope in the
visible region.

The source of ultraviolet radiation may be a resonance (cold
cathode) or other type mercury arc, or a metallic spark (*e.g.*, with
rotating cadmium electrodes).   Since glass is opaque to short wave-

length radiation the optics of the microscope and the illuminating apparatus must be of quartz or other transparent material. Liquid filters, Christiansen filters, or a monochromator are used to isolate the desired wavelength. Much of the work has been done at 257 and 275 mμ with monochromatic objectives (20,21), although both longer and shorter wavelengths have also been used.

Since the eye is insensitive to these wavelengths, focusing of the microscope must be done indirectly and the results observed from photographs or from energy measurements on isolated parts of the image with a photocell or other sensitive instrument. Originally a fluorescent finder-ocular was used for focusing, or else a number of pictures were taken by trial and error. The recent improvement of fluorescent screens makes possible direct focusing in a darkened room when the observer's eyes are dark-adapted, thus insuring good photomicrographs with a minimum of difficulty (22). The mounting medium, slides, and cover glasses must be transparent to ultraviolet. Living tissues are damaged by short wavelength ultraviolet so that exposures must be kept to the minimum. The eyes of the observer must be protected from the radiation or a severe conjunctivitis will result.

Crystals just smaller than can be seen with light have been photographed for measurement and study of their form with the longer wavelength radiation at 365 mμ. Since glass is transparent to this radiation the ordinary microscope may be used (23).

The catoptric microscope has also been used for ultraviolet microscopy (Sect. L2) and Brumberg (19) has described the use of a reflecting objective, with revolving filter discs, fluorescent screen, and means to convert the ultraviolet into visible light so that the object is seen in color according to the position and nature of its ultraviolet-absorbing bands.

The ultraviolet microscope has been used to examine fine crystalloids, silver halide grains, nuclear detail in tissue cells with relation to the question of genes and enzyme formation, location of proteins and their concentration changes, regenerating nerve fibers, muscular dystrophy, bacteria and virus particles, latex, rubber, and emulsions. Some of the work is done with direct photographs, but the present tendency is to use the instrument as a spectrophotometer to actually measure the absorption bands of the specimen. Semiapochromatic objectives are being designed to focus simultaneously several wave-

lengths of ultraviolet. Since monochromators are not convenient sources, research continues toward the improvement of filters for isolating the necessary regions of the spectrum.

Even shorter wavelength radiation could be used for absorption spectra reasons or to attain still better resolution. Between 200 mμ and the wavelength region where image formation fails to take place, however, the radiation is absorbed very strongly, even by air. This necessitates working in a vacuum; also special photographic emulsions must be used. Little exploration has been done in this region and its possible advantages are unknown, but the difficulties are rather formidable. A simple X-ray microscope has been proposed by Hámos and Engström (24).

This is a specialized field that requires a good knowledge not only of microscopy but of radiant energy measurements. Ultraviolet investigations should not be undertaken unless one is prepared to master a complex and time-consuming technique. The research program should be planned in advance of procuring equipment.

D. OPAQUE SPECIMENS

1. Direct Illumination

Specimens too dense to transmit any light must be illuminated from above. When low magnifications are adequate, light can be reflected onto the specimen with a mirror or prism, or directed onto the specimen from a small lamp having a focusable lens system to concentrate the light where it is most useful.

The microscopist of the previous century used a Lieberkühn mirror surrounding the objective to reflect light from the substage mirror onto the top of the specimen. The Silverman Illuminator added a Lumiline lamp to the reflector and was very useful. Another helpful source has a ring of small lamps that surrounds the objective (58). A method that is useful for photomicrography of small objects is to wrap some translucent paper in the shape of a cone around the objective; by directing the light from one or more lamps onto the outside of the cone, a diffused light falls on the specimen and harsh shadows are avoided. In the case of higher power objectives that focus so close to the specimens, the methods of Sections D3 and D4 become necessary.

2. The Stereoscopic, Biobjective Microscope

Opaque specimens may be examined to advantage with the stereo-scopic, biobjective binocular microscope, although it is not limited to such specimens (*cf.* Sect. J1).   The microscope (Fig. 3) is really com-posed of two microscopes, one for each eye.   Erecting prisms are
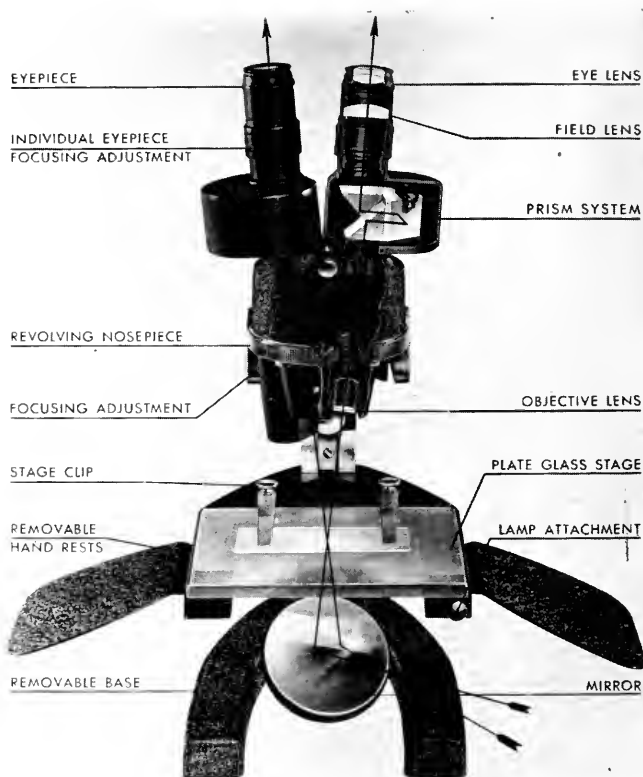
Fig. 3.   The biobjective, stereoscopic microscope.

included so that the image is not inverted, as with the monobjective compound microscope, and dissecting operations and other manipula-tions made on it are normal.   Since each eye sees a separate view from a different angle the microscope shows real stereoscopic depth and the greatest applications concern observations in which the third dimension is essential.   When the converging angles between the

objectives and the oculars are the same and the eyes are placed at the
unit plane of the instrument, the image is orthostereoscopic.   How-
ever, microscopists often prefer to have the depth emphasized and
the convergence of the objectives is increased to give hyperstereos-
copy.   While in general this is helpful, some distortion does occur
and should be considered in interpreting the image.

The limitation of the instrument arises from the mechanical dif-
ficulty of placing the front lenses of the two objectives close together,
resulting in the numerical aperture's being limited to about 0.12.
With the working rule that for the average eye the useful magnifica-
tion is 1000 N.A., there is no need for greater magnifications than
about 120 to 150×.   In fact they are most useful for specimens seen
at 5 to 30× magnification, because at the lower magnifications more
of the specimen can be seen at one time.

The instruments have means for moving the oculars to accommo-
date the separation of the observer's eyes; also, one eyepiece is usu-
ally separately focusable for individuals with unequal eyes.   The in-
strument is focused with a simple rack and pinion since a fine adjust-
ment is unnecessary at these low magnifications.   No special skill is
required of the user.   It is well to check the instrument occasionally
by looking at a small ruler placed horizontally and then vertically
across the field of view.   The edges of the field for each eye (seen when
the other is closed) should be the same.   If not, the instrument
should be corrected by the maker or by a competent person.   With a
comfortable viewing position and adequate, but not too bright, il-
lumination the microscope can be used for considerable periods of
time without strain.

3. Vertical Illumination

The vertical illuminator (Fig. 4) uses a half-silvered cube or a
coated thin, flat piece of glass mounted at the back of the objective
to reflect light, from a source at the side, through the objective to the
specimen.   Light returning from the specimen passes through the re-
flector to the eye.   A totally reflecting prism, small enough to cover
only part of the aperture, has also been used.   Vertical illuminators
are useful with polished metal surfaces and other specimens of high
reflectivity.   Some glare results since the light goes through the ob-
jective in both directions.   For best observation antireflection coat-
ings are essential for all glass–air surfaces to reduce the glare.   When

EYEPIECE

PATH OF LIGHT

VERTICAL ILLUMINATOR

COATED
REFLECTOR

DIFFUSING
FILTERS   DISKS   BULB

QUICK
CHANGE
NOSEPIECE

OBJECTIVE
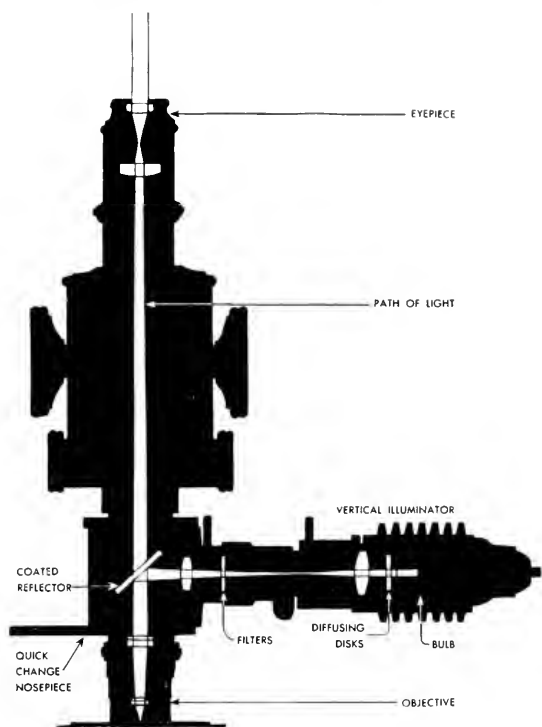
Fig. 4.   Light path through microscope with vertical illuminator.

Light from specimen
to ocular

Objective
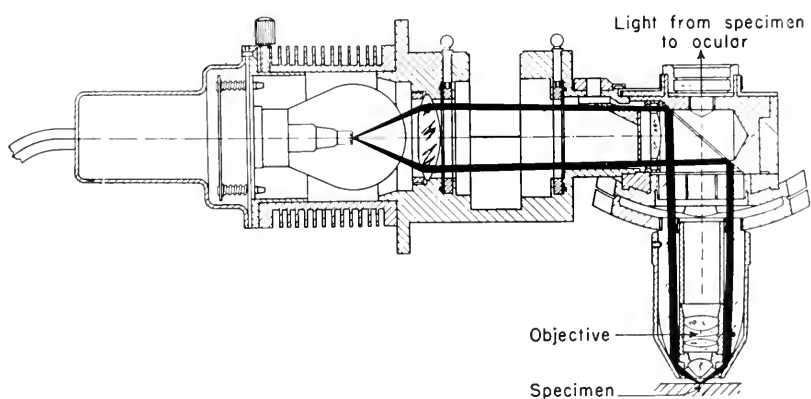
Specimen

Fig. 5.   Light path for epi-illumination vertical illuminator. (Courtesy *Bausch
& Lomb Optical Co.*)

the surface of the specimen polarizes light, or introduces optical path differences, polarization or phase equipment assists in obtaining the best contrast and detail in the image. The vertical illuminator is of use only with specimens that have adequate reflecting properties and has attained greatest use in metallurgical microscopy (*1,3,4*).

4. Epi-illumination

The epi-illuminator has a condenser built around the objective, which concentrates the light onto the specimen so that only the light reflected from the specimen goes through the objective (Fig. 5). Thus much less glare is involved and specimens with lower reflecting power can be examined. Special, slender objectives are required and these may be obtained for all magnifications. For a range of magnifications, however, several condensers may be required. Some of the units are small enough to be used on a revolving nosepiece. Transparent specimens are examined on a black glass or other opaque reflecting slide. Insect appendages, tissue circulation, and microorganisms have been studied in this way. When the optical apparatus is made of materials transparent to ultraviolet, the equipment can be used for fluorescence and ultraviolet microscopy. Some skill is required and the method has not yet come into general use (*58,65, 66,68*).

E. SMALL AND ULTRAMICROSCOPIC SPECIMENS THAT SCATTER LIGHT

1. Darkfield Microscopy

Light is concentrated at the specimen by the darkfield condenser in the form of a hollow cone at such an angle that none enters the objective unless a specimen is present to change its direction (Figs. 6D, 7A). Small particles that scatter light are readily seen and materials with large optical path differences show better with darkfield than those with small path differences.

Two types of darkfield condenser are in general use, the bispheric and the cardioid. Both have two spherical reflecting surfaces to control the reflection and concentration of light. A nonspherical cardioid shape for the outer reflector of the latter type would give better concentration of light at the specimen, but the difficulties of grinding such a shape have so far prevented commercial use of the

design. A simpler, less efficient construction uses a paraboloid mirror and is useful for less critical microscopy. For elementary work a center stop having an opaque region of proper diameter (*ca.* 17 mm.) can be inserted into the lower part of the microscope condenser.

A   B

C   D
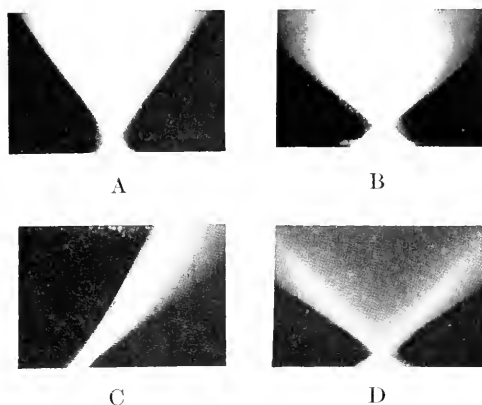
Fig. 6. Comparative cones of light: condenser without (A) and with (B) immersion oil contact to the slide; (C) oblique lighting; (D) darkfield lighting.
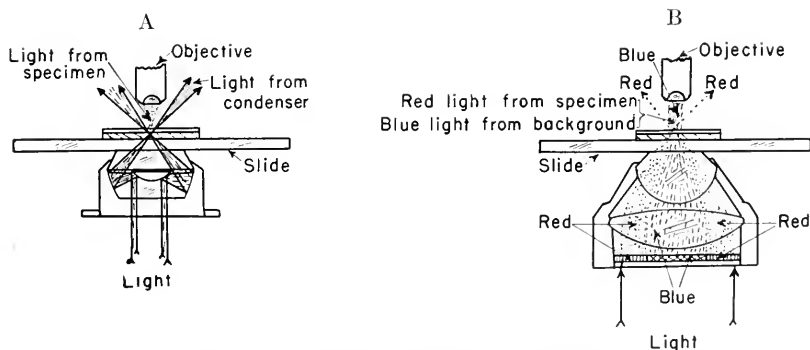
Fig. 7. Light path through bispheric, darkfield condenser (A) and through condenser with optical staining disc (B).

More elaborate condensers have means for focusing the light and even changing from bright- to darkfield (59). The darkfield condenser must have an immersion fluid contact with the under surface of the slide. With high aperture objectives, it is often necessary to

insert a stop near the back lens of the objective to decrease its aperture to match that of the condenser.

Parallel light should be used and the condenser must be carefully centered for uniform illumination. The slides must be very clean and free from scratches and imperfections and of the thickness for which the condenser is designed. Colored or stained tissues may be examined. Particles smaller than the limit of resolution may be detected when the light is strong enough, although their shape and size cannot be measured. In fact, the smallness of the particles that may be detected depends only on the intensity of the light. Very strong illuminants should be used to make the energy in the center of the diffraction disc (see Fig. 1) great enough to stimulate the retina and be seen. Some information as to the number of particles and their average size may be inferred when the depth of the specimen and area examined are known. The darkfield microscope is useful for locating the spirochetes of syphilis in exudates from sores and in the study of small cellular inclusions and the larger colloidal materials. Microscopes with built-in, precentered darkfield condensers are available (57) and most brightfield microscopes can be adapted for darkfield microscopy by substituting a darkfield condenser in place of the brightfield condenser. Some practice in centering the equipment and interpreting the image is required, but the method is useful and not particularly difficult (4).

2. Ultramicroscopy

The ultramicroscope passes a narrow beam of intense light through the specimen from one side and the observer sees the specimen by its Tyndall effect, as with particles in a beam of sunlight. It is useful for the study of smokes and colloids and requires special equipment and considerable skill for efficient use (3).

3. Optical Staining

When the central part of the condenser stop is made of a transparent colored material (rather than opaque as in darkfield stops) and the annulus surrounding it is made of a complementary colored material (Fig. 7B), reasonably transparent specimens will appear as if stained by the color of the annulus against a background of the color of the center stop. Crystals, inclusions, and fairly transparent microscopic plants and animals may be observed in this manner. Col-

ored materials have been examined to show the penetration of the dye and the nature of the parts of the specimen stained.   The stops are available as discs with various color combinations (62), with a metal form to hold interchangeable colored discs and annuli (58), and as a complete substage condenser unit (68).   The results are spectacular, but often do not show as much detail as darkfield illumination. Royer and associates have recently evaluated the method (25).

Optical staining may also be obtained by utilizing the differences in dispersion of light by regions of the specimen having different refractive indices for colored light of different wavelengths.   The specimen is mounted in a medium of proper refractive index and illuminated with white light by means of a darkfield condenser.   Regions of different index, or bending power, are then seen in different colors. Differentiation within tissues and minerals is possible and the method is useful for locating impurities in pulverized material.   The method is new, although based on known principles, and promises greater utility with low and medium than with the greatest magnifications (26).   Besides a darkfield condenser, media compatible with the specimen must be available in a series with small differences in refractive index, or must be made by mixing as required to bring out the color differences of the specimen.   Both methods permit examination. of many materials with little or no damage to the specimens.

F. SPECIMENS THAT POLARIZE OR CHANGE POLARIZED LIGHT

When the direction of vibration of light changes or when the speed of the light depends on its direction of vibration as it passes through the specimen, polarization microscopy will be useful.   Ordinary light consists of electromagnetic vibrations in all planes at right angles to the direction of its travel.   Some substances polarize the light that passes through them so that the emergent beam vibrates in a single plane, or other geometrical form.

Special optical equipment must be added to the ordinary microscope in order to permit the measurement and analysis of the various changes in light that may be produced by such an object.   A polarizing device, called a polarizer, is placed below the microscope condenser and another (the analyzer) is placed over the ocular, or within the body tube of the microscope;   at least one of them should be provided with a scale for measuring angular rotation.   The polariz-

ing material may consist of calcite prisms or a special grade of Polaroid.   Objectives must be obtained that are free from strain.   A rotating mechanical stage is desirable.
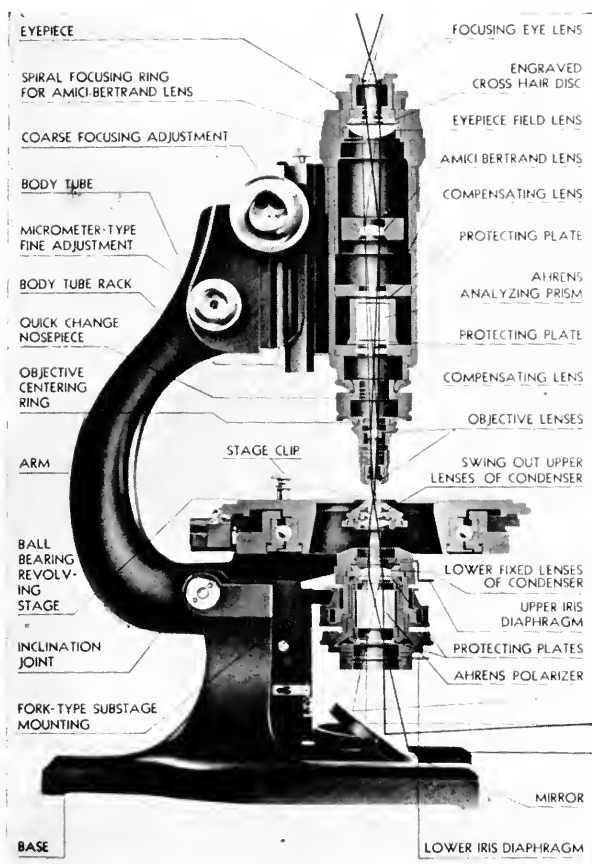
Fig. 8.   Polarizing microscope showing parts
and the path of light through the instrument.

The above minimum equipment may be added quite readily to an ordinary compound microscope for polarization study.   For critical and analytical work a standard polarizing microscope (Fig. 8) is preferable.   In addition to a Bertrand lens the latter has provision for the insertion of retardation plates, also a rotating stage with calibrated scales.

The polarization microscope is used to identify materials, especially minerals. Materials that possess a single, refractive index are called isotropic and do not appear different in polarized light. Anisotropic crystals have more than one refractive index and are characterized by having one or more optic axes. From the observation and measurement of their axes, refractive indices, and crystal habit, microscopic specimens may be identified by reference to tables of such data (*3,27a,30a*).

Birefringent objects, and materials that become so under strain, can be studied by the polarization colors they form, either with or without retardation plates, quartz wedges, and other compensators. The changes in birefringence with function, as the contraction of muscle fibrils, may also be studied in this way. The membrane surrounding red blood cells has been measured with the aid of a special comparison polarization microscope (*27b*). Liquids that contain asymmetrical particles become birefringent when forced to flow through a capillary tube, or to move so as to orient the particles. Solutions of proteins, myosin, tobacco mosaic virus, hemocyanins, fibrinogen, and other materials have been studied with the polarizing microscope (*28*). Skeletal elements of many organisms, fibers, hairs, and many tissues of plants and animals are birefringent and the polarization microscope is useful for their identification and analysis (*3, 6* (3rd ed.), *29, 30*).

A qualitative study of the images formed by anisotropic or birefringent specimens requires only a minimum of equipment and training, but the use of a polarizing microscope as an analytical tool requires considerable experience and skill.

G. FLUORESCENT AND PHOSPHORESCENT SPECIMENS

Many materials absorb one wavelength of radiation and reradiate the energy at another, usually longer, wavelength. Such specimens, when irradiated with invisible ultraviolet, may be seen by means of the visible fluorescence emitted. Other materials may be treated with selectively absorbable fluorescent chemicals, which is analogous to a staining procedure, or a combination of fluorescent chemicals and nonfluorescent dyes may be used (*31a*). Proteins and many materials of natural occurrence are fluorescent. Great contrast may be obtained with bacteria so that they appear bright and self-luminous against a dark background, *e.g.*, the identification of tuberculosis and

other acid-fast bacteria (*31*, p. 452), and for detecting important chemicals like riboflavin and vitamin A. Some observers believe malignant tissues may be identified with the fluorescence microscope (*32*).

Since long wavelength ultraviolet is transmitted by optical glass down to about 310 mμ an ordinary microscope may be used. For work with shorter wavelength radiation a quartz or Corex condenser is required. The carbon arc is the strongest ultraviolet source, although for many studies the mercury arc or the tungsten lamp may be used. Filters are used to remove the visible light.

Two methods are used. One method employs a darkfield condenser so that none of the ultraviolet can pass directly into the objective. The resulting fluorescence usually is less bright; the observer's eye must be well dark-adapted and the observations made in a darkened room. Photographic exposures may require hours.

The other method uses the crossed filter technique. A filter that transmits only the radiation to be absorbed (*e.g.*, blue or ultraviolet) is placed over the source; the eyepiece filter (*e.g.*, yellow filter) does not transmit this direct radiation from the source, but will let through the fluorescent light. This procedure gives brighter images and makes possible shorter exposures. When the fluorescence involves all colors, the filter on the source must not transmit any visual light. Since some ultraviolet will pass through the microscope, an ultraviolet-absorbing filter should be used to protect the eye and is required when photographs are made because many emulsions are more sensitive to short wavelength radiation. The lens of the eye fluoresces and obscures vision when such a protective filter is not used. Fluorite fluoresces so that apochromatic and fluorite objectives are less satisfactory than achromatic objectives. The mounting medium and the immersion oil used must be nonfluorescing, which disqualifies cedar oil, balsam, and many of the usual materials that fluoresce strongly. Glycerin is a good mounting medium and some of the modern synthetic immersion oils are suitable.

The need for a darkened room and dark-adapted eyes restricts fluorescence microscopy and the observations may be time-consuming, since some of the phenomena are transient. The image is not too bright and longer photographic exposures are required. The method offers a promising field for the investigator desiring to work in a realm with few known rules and guideposts and relatively unknown possibilities (*31–33*).

When the light is emitted for a short period after the exciting radiation is turned off the material is called phosphorescent; a microscope for the examination of such specimens has been designed by Harvey and Chase (*34*).

II. SURFACE IRREGULARITIES AND OPTICAL NONUNIFORMITY OF SPECIMENS—INTERFERENCE MICROSCOPY

Two plane-parallel pieces of glass in perfect contact will show only a uniform color when monochromatic light is passed through them, but, when the surfaces are not in uniform contact, interference fringes are seen around any imperfect region; when the two surfaces are inclined to each other interference bands become visible. The fringes are due to the combination of the light waves from the multiple reflections (*38*). Brighter fringes are seen when the two reflecting surfaces are covered with a partially transparent film of silver, platinum, or other good reflector. Interference methods are useful for the measurement of surface irregularities, small displacements, and inhomogeneities in transparent material and are generally used for precise measurements in the testing of optical elements and with gage blocks. A special interference microscope has been made for the examination of metal and other highly reflecting surfaces (*68*). Microscope fine adjustments are calibrated with an interferometer (*1*; *cf.* Sect. L3).

A multiple beam method has been used by Tolansky (*35*) for the examination of crystal surfaces and extended to biological surfaces by Greenham (*36*), who pictures the cuticular surface of an apple leaf. Merton (*37*) reported that it was unnecessary to use expensive optical flats, since selected pieces of glass or plastic films could be partially platinized and used. The specimen is mounted between two pieces of platinized glass, or plastic and glass, and examined with a brightfield microscope using monochromatic light. The distribution and shape of the interference fringes indicate the structure of the specimen. The optical path is $2\,tn\,\cos\theta$ where t is the distance between the plates, n is the refractive index, and θ is the angle made by the rays with the normal. Thickness is obtained by counting the number of fringes; N fringes denote a thickness of $N\lambda/2$. Broader fringes were obtained when the lower side of the support plate was a ground surface that revealed considerable detail within the transparent specimens.

A brightfield microscope without the condenser is used with monochromatic light from a sodium or mercury arc (screened with proper filters to isolate the line desired), or tungsten light with an interference filter (*e.g.*, 555 mμ) (*63*), for the examination of the specimen mounted between suitable plates with partially metallized surfaces. Merton states that for practical purposes the method is limited to a specimen a few microns thick, but this will actually depend on the transparency and optical path differences in the specimen. Surface details and internal detail due to optical path differences may be visualized with the interference microscope, but it has been too little used with biological materials to suggest what will be its ultimate contributions and limitations.

I. TRANSPARENT AND SLIGHTLY ABSORBING SPECIMENS—PHASE MICROSCOPY

Phase microscopy utilizes optical path and absorption differences in the microscope to increase or decrease, or reverse and increase or decrease, the contrast in the image from transparent specimens having optical path differences, or of low absorption contrast (Fig. 9).
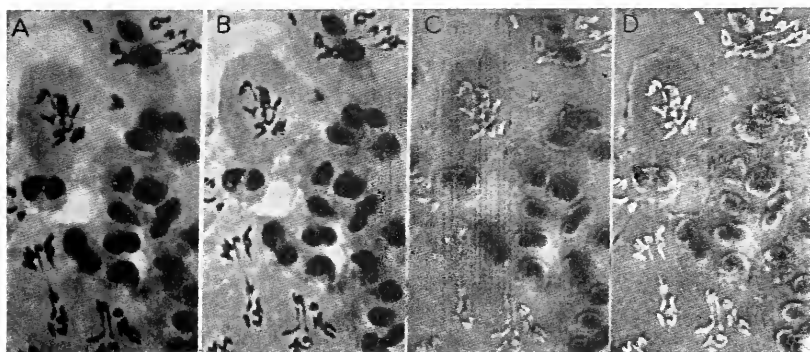
Fig. 9. Wheat chromosomes stained with acetocarmine: (A) brightfield microscope; (B) dark contrast; (C) bright contrast; (D) bright contrast plus red filter with phase microscope.

The method is applicable to living unstained cells and tissues, crystals, colloids, natural and artificial fibers, surface and internal detail of glass and plastics, minerals, stained or colored materials of low contrast, and replicas of surfaces (*39,40*).

Light waves passing through an optically denser medium as at C in Figure 10 are slowed with respect to those not passing through the denser medium, but the difference does not affect the eye or photographic plate. An absorption difference such as E does not alter the phase relations but changes the amplitude and is visible. An annular diaphragm, K, is placed in the condenser of the phase microscope and an image of its illuminated opening is formed at the back focal plane, J', of the objective. At this plane is placed a diffraction plate which may consist of either (1) a thin layer of metal and/or dielectric, as at F, of proper size to cover this image of K (conjugate area), or (2) a complementary area of similar composition as depicted at I,
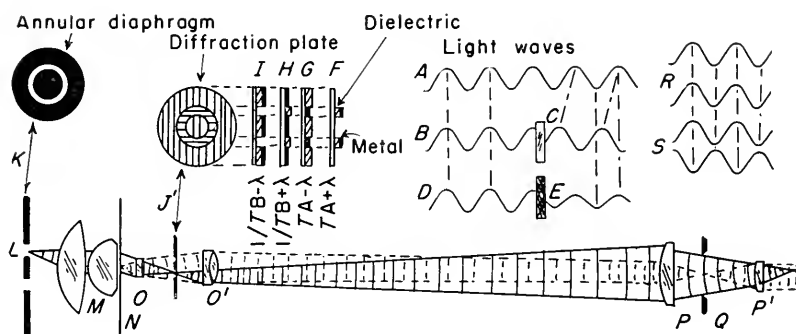
Fig. 10.   Diagram of the phase microscope (57).

or (3) some desired combinations of the I and F type distributions as shown at G and II.   When a specimen is placed on the phase microscope it diffracts light; this deviated light passes through the entire objective, while the undeviated light must pass only through the conjugate area.   The diffraction plate thus has a differential effect on the light.   When the image is formed by the combination of all the separate rays from the specimen and its background, light vibrations of the same phase and amplitude, R, add to give increased brightness; likewise those of opposite phase and amplitude, S, produce darkness.   Intermediate phases and amplitudes combine to give grays. Diffraction plates of the A+ and B+ types give bright contrast, as regions of greater optical path appear brighter than those of lesser path and conversely, the A— and B— plates show the regions of greater path as darker. By changing plates having different thicknesses of absorbing metal and/or retarding dielectric layers, the image contrast may be changed to optimum for any given specimen.   Further explanation of the theory requires mathematical analysis, which is available elsewhere together with a history of the method (39).

The microscope requires no special skill for its use. An auxiliary telescope is used in place of the eyepiece while viewing the diffraction plate in the objective for the purpose of centering the microscope condenser and the annular diaphragm. On replacing the eyepiece the instrument is ready for use. A microscope illuminator with a focusing lens system is desirable and for greatest contrast monochromatic light is preferable. For many materials and living cells good bright contrast is obtained with a $0.2A + 0.25\lambda$ diffraction plate and dark contrast with a $0.2A - 0.25\lambda$ plate. Increased contrast can be obtained with the $0.14A + 0.25\lambda$ and $0.14A - 0.25\lambda$ plates and greater contrast obtained with lower transmission plates of $0.07A \pm 0.25\lambda$, especially with very small specimens. The $1B - 0.25\lambda$ diffraction plate is useful with supravitally stained preparations, Feulgen stains, and with colored or absorbing material of low contrast. Sudan-black-stained material shows with the greater dark contrast $2.5B - 0.25\lambda$ plate, as do plant chromosomes and yeast cells. For special applications plates of other than a quarter wavelength retardation give better contrast (57). Some specimens show better with bright or dark contrast, some equally well with either and some show detail with one not seen with the other. Likewise some observers may prefer to use bright or dark contrast. Dark contrast more nearly resembles stained material and is helpful for measuring. Bright contrast seems better for counting. Bright contrast will reveal the presence of smaller objects when the intensity of the light is adequate and is useful for stereophotomicrographs (40b). Color contrast may be added to phase contrast (Fig. 9). Some experience with the equipment is necessary for its efficient use and a proper understanding of the image.

Specimens with small optical path differences are better than those with large optical path differences for phase microscopy. In fact it is often possible to improve the contrast by varying the optical path with media of slightly different refractive indices. Specimens with large path differences or with good color contrast are better examined with other microscopic methods. The phase microscope has made possible the direct examination of both transparent and low absorption contrast specimens that hitherto could be examined only by staining and other indirect methods. Even though it has been used only for a short period of time, advances are being reported in biology, medicine, and industry (38–40,57,58,60,68).

J. SPECIMENS THAT REQUIRE MANIPULATION

1. Microdissection, Microincineration, and Radioautographs

Equipment for manipulation of materials under the various powers of the microscope varies from a simple mechanically controlled tool mounted on a base with three mutually perpendicular rack and pinion movements to elaborate controls for needles, knives, or pipets. Some work on a pantograph principle and others have hydraulic controls. The equipment may be adjustable so that the full movement of a control may include only the diameter of the field of view at the magnification used. The equipment is useful for dissection, exploration of structure, injecting or placing materials and reagents, and still other procedures possible only with high magnifications. Special chambers are required to maintain the proper humidity and temperature for work with living tissues. The simplest equipment may be homemade and require little skill and practice for isolating bacteria, yeasts, and spores (*41*). More elaborate equipment requires considerable skill and experience for its efficient use as well as in making the microneedles and other tools used with it (*6*, p. 62). Microincineration of tissues on quartz or hard glass slides is possible in a small furnace and the location and identification of mineral components may be determined with chemical and other micromethods. Owing to the fragility of the specimens considerable skill and care are required of the microscopist (*4,42*).

The location of radioactive materials and tagged chemicals in tissues (see also Chapter XVII) may be accomplished by placing the tissue section on a photographic emulsion long enough for the particles, or rays, to affect the plate, developing the plate, and correlating the traces in the emulsion with their origins in the cells of the tissue. The phase and darkfield microscopes offer promise in this new and difficult field of microscopy (*43*). Useful techniques should include stereophotomicrographs for the lower magnifications with sufficient depth of fields, or divided, double exposures on the same film with part of the exposure at the level of the tracks and part at the level of the tissue. Nonstereoscopic pairs of tracks and of tissue can be superimposed with the stereoscope to aid in locating the origin of the tracks.

The last two methods require that the microscopist be skilled in preparing and sectioning the material as well as in its interpretation.

2. Stereoscopic and Inverted Microscopes

The stereoscopic, biobjective microscope described in Section D2 is especially well adapted for manipulation of materials for examination at magnifications up to about 50×. Although these instruments are customarily furnished with the usual stand and a large glass stage, they may also be used either directly on the table by removing the base, or mounted on a long arm attached to a pedestal, or mounted on a base board.

For the study and microdissection of tissue cultures and for large, heavy specimens an inverted microscope may be more convenient (Fig. 11). The specimen is placed on the stage and its underside ex-
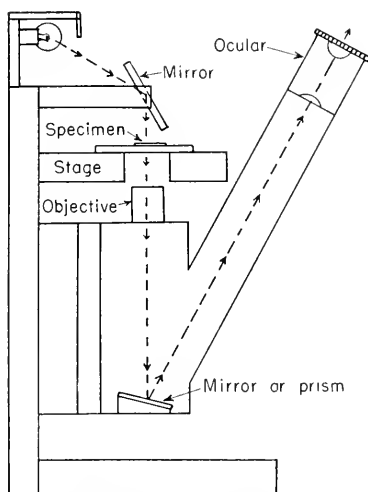
Fig. 11.   Diagram of an inverted microscope.

amined by the objective, with a prism at the base to reflect the light from the specimen up through the ocular to the observer. For transparent specimens a mirror and condenser are mounted above the specimen to reflect light through it to the objective. Special metallographic microscopes are available for the examination of metal surfaces and include a light source with vertical illuminator, inverted microscope, and camera.

K. INACCESSIBLE SPECIMENS AND SPECIALIZED MICROSCOPES

The laboratory type and the more generally used special microscopes have been described. There are other kinds that should be mentioned with reference to their adaptations to special problems. Folding pocket microscopes, for example, are convenient for field work and vary in size and equipment from a simple pocket instrument of limited magnification to folding microscopes with a nosepiece and three objectives that fit into small carrying cases.

Many kinds of microscopes have been built for such special uses as the observation of the interior of body cavities. Some endoscopic instruments with low magnifying power have prisms for seeing around corners, *e.g.*, gastroscopes, laryngoscopes, etc. Capillary microscopes are used to examine changes in the capillaries of the finger (*65,68*). A cathetometer consists of a microscope mounted on a vertical moving member and provided with a scale and vernier for measuring small distances, *e.g.*, of growing organisms (*11a*). Other mountings move the microscope body horizontally and are used for measuring spectrograms, verifying scales, and measuring short distances (*64*). Some microscopes have elaborate, calibrated stage mechanisms for holding specimens and turning them for the measurement of angles or distances as in the Toolmaker microscope (*64*). Universal microscopes include built-in light sources and various special microscope bodies may be used on them interchangeably (*65,68*).

Microprojectors consist of a microscope and light source for projecting the image onto a screen. The smaller ones have tungsten lights and the larger ones are provided with intense carbon or mercury arcs. Industrial microprojectors usually have a self-contained screen and may have scales for use on the image as in contour projectors.

For work at high magnification it is necessary to design the equipment for freedom from vibration and it is desirable to have better control with slower focusing motions. Notable designs for horizontal microscopes have been described by Lucas (*44*) and by Graton and Dane (*45*). Harvey (*46*) designed a combined centrifuge and microscope to observe the specimen during centrifugation.

L. USEFUL ACCESSORY EQUIPMENT AND RECORDING METHODS

1. Illumination and Its Importance

No matter how excellent the microscope, detail can be seen only when the illumination is equally excellent. What may be seen with the microscope, true or false, depends on proper lighting. Daylight is too uncertain for microscopy and has been replaced by electric illuminators. These may be divided into two classes: (1) illuminated surfaces and (2) lens systems for concentrating the light. With the former, the lamp, mirror, and condenser of the microscope are arranged so that both the field of view and the back lens of the objective, seen on removing the ocular, are uniformly filled with light. If the microscope has no condenser, one uses the concave mirror and moves the lamp to the distance from the microscope that fills the field with light.

Lamps with an iris diaphragm and lens system are essential for the best illumination of the microscope for comfortable vision and to fill the aperture of the objective for full resolution without glare. With the coiled-coil filament lamps (e.g., CC-13 filament), the lamp condenser is focused to give an image of the filament coils on the iris of the microscope condenser; then the microscope condenser is adjusted to bring the diaphragm of the lamp condenser into focus on the specimen and finally the microscope condenser aperture is opened until the back lens of the objective is just filled with light. This method of adjustment often is called Köhler's method.

When the light source is large and sufficiently uniform to fill both the field and the back aperture of the objective with light, the lamp condenser may be focused to infinity (for practical purposes the filament is focused on a wall ten or so feet away) and the microscope condenser adjusted to focus the filament on the specimen. This method is called "critical" illumination, although it is actually no more critical than Köhler's and other methods (10,47). This method should be used for darkfield and fluorescence microscopy and for optical staining.

Tungsten filament lamps are good light sources. The ribbon filament type gives the most uniform illumination. Since the entire ribbon may not be imaged on the condenser and usually does not operate at as high a color temperature, it actually gives less light than the coiled-coil type filaments. Of the latter, more of the CC-13 filament

can be imaged in the microscope than of any other shaped filaments so that its illuminating efficiency is higher. Lamps of the same type of filament with longer lives burn at lower temperatures and give both yellower and less light than those burning at higher temperatures with correspondingly shorter lives.

The carbon arc is the brightest artificial source available and suitable for microscopy. Its disadvantages stem from the frequent carbon replacement and a certain amount of necessary continuous adjustment. With automatic line voltage control and dry carbon arcs much of this difficulty may be avoided. The zirconium arc is less bright than the carbon arc and brighter than tungsten filament lamps. The light from it is very uniform but special lenses are necessary to magnify its small area and special equipment is required to control the electrical supply. Heat-absorbing glass should be placed between an intense source and the microscope to protect living organisms.

2. Accessory and Special Equipment

A back surface plane mirror gives two images, a weak one from the front glass surface and a strong one from the silvered surface. When this double image is objectionable in critical microscopy, it can be avoided by using a prism, or a front-surfaced mirror, mounted in place of the usual mirror.

The concave mirror, used without a condenser, does not give a wide enough cone of light to illuminate fully the aperture of other than low power objectives. A condenser and the plane mirror should be used with all objectives having a numerical aperture greater than 0.25. Unless the aperture of the objective is filled with light it will not resolve the optimum detail otherwise possible. The better condensers are corrected for aplanatism (coma and spherical aberration) and for chromatic aberration. Corrected condensers should be provided with centering mounts, unless these are built into the microscope substage.

Dry objectives of large numerical aperture are sensitive to spherical aberration arising from cover glasses of other than the standard 0.18 mm. thickness. They should have a correcting collar or be used with an adjustable draw tube, or with a tube length corrector. Immersion lenses are less sensitive to cover glass thickness because of the homogeneous oil contact, but the dispersion of the cover glass may

affect the image and it is preferable in critical microscopy to correct
for any residual aberration.

When a specimen is mounted in a denser medium than air, a
medium of the same refractive index should be used both between
the condenser and the slide and between the cover glass and the ob-
jective; otherwise, the numerical aperture may be limited to unity
with a corresponding restriction on the resolution. For objectives of
greater aperture, the immersion fluid for which they were designed
(water, glycerin, or an oil of the same index as the glass) must be
used (Fig. 6A–B).

During the early nineteenth century catoptric, or reflecting, ob-
jectives were preferred to those made with lenses. Curved mirrors
were used in place of lenses. At that time achromatic lenses were
not yet available and microscopists probably could do better with the
reflecting objectives. Mirrors do not have chromatic aberration and
so are in focus for infrared light and for ultraviolet. The use of mir-
rors of aspheric shape avoids spherical aberration and with a suit-
able metal surface short wavelength radiation may be used with a
corresponding gain in resolution. Some light loss occurs at the open-
ing in the mirror to pass the light on to the ocular, or from the second
mirror used in the system. At present there is a return in interest to
the possibilities of reflecting microscopes (19,48–51).

The binocular body is more comfortable, since the work of seeing
is shared by both eyes, but does not usually give a stereoscopic image,
because each eye sees the same identical image. A stereoscopic ef-
fect of depth may be obtained (1) by setting the interpupillary dis-
tance so that the iris of each eye cuts off half of the light from the ocu-
lar, (2) by placing stops within the prism system to accomplish the
same effect, or (3) by using D-shaped stops over the eye points (Rams-
den discs) of the oculars. Other methods involve placing a prism over
a lateral half of the back aperture of the objective, so arranged that
the light from each half goes to opposite eyes, or by a Polaroid disc
with lateral halves having polarizing axes at 90° to each other and
Polaroid eyepiece filters adjusted so that each eye sees the light com-
ing from one-half of the objective (58). The half-aperture methods
reduce the resolving power by one-half, which may or may not be
significant, depending on the fineness of detail being observed.

Huygenian oculars give a bright image that has distortion at the
edge of the field and should not be used for measuring and drawing of
detail much beyond the center of the field. Compensating oculars

and positive oculars are practically distortion-free, but often have too much curvature of field for photomicrography. Special oculars are available and should be used for projection, photography, and special applications. A comparison ocular with a split field may be placed over two microscopes when close comparisons are required. The demonstration ocular makes simultaneous observation possible for two observers.

Microspectroscopes have been designed for use with the microscope (*3,4,52*) and microcolorimetry and refractometry are also possible (*3,4,53*).

Once a detail is located on a slide, it may be important to mark its place so that it may be found another time. This can be done crudely with a drop of ink just below the object, but better by recording the coordinates of a graduated mechanical stage, or by replacing the object slide with a Maltwood Finder or other slide having numbered squares on it and then recording its position number; reversing the process leads one back to the same area on the specimen. Other useful methods mark the cover glass over the specimen (*54*).

3. Counting and Measuring

A ruled scale placed in the eyepiece with the ruled side next to the top of the diaphragm is seen superimposed on the image of the specimen and may be used for counting or measuring. The value of the scale unit depends on the magnification of the ocular and objective and is obtained by calibration with a stage micrometer. Any change of the magnification resulting from substituted optics, readjustment of the draw tube of the microscope, or of the interpupillary distance when inclined ocular tubes are present will change the unit value and require recalibration. With an adjustable draw tube the distance subtended by the reticule scale can be changed, within limits, to give convenient integral values. Since the spherical aberration is changed with the change in tube length, this method of adjustment should not be used for critical purposes.

Filar, or screw eyepieces may be obtained with movable scales for precision measuring. Measurement can be made also with a mechanical stage with graduated scales. A special integrating mechanical stage with several movements may be used for estimating percentage composition of a specimen by means of area and/or volume determinations.

Fine adjustments of microscopes are calibrated for measuring vertical distances, but limitations arise in the depth of focus of the objective and the precision of the fine adjustment. To avoid hysteresis and other errors in the mechanism, the measurement must always be made in the same direction, *i.e.*, either up or down. When precision measurement is required, the fine adjustment should be calibrated with an interferometer (*1*, p. 194). The additional use of long levers and protractors to give finer readings or movements may be misleading in that they can exceed the precision of the fine adjustment. When the specimen is in a medium of different refractive index than air the apparent thickness must be multiplied by the index of the medium to get the true thickness.

When a suspension containing particles is placed in a cell of known thickness, counts per unit volume may be made from knowing the area of the field of view or with a net ocular reticule. The Whipple disc is useful both for counts and for estimating percentage composition. Another method uses a chamber of known depth that has a scale ruled on its base as in the familiar hemocytometer used for blood counting. Various rulings are available and the chambers may be purchased with various depths. Counting and measuring require care, careful calibration, and proper enumeration and evaluation of possible errors. Studies on the errors of blood counting are useful in suggesting appropriate procedures for other types of counts (*55*).

4. Drawing, Projection, and Records

Good records are a necessary part of any research and those from observations with the microscope may be made by drawing or photography. A prism or a mirror can be mounted over the ocular, the image projected onto a drawing surface and the outline filled in with as much detail as desired. The "camera lucida" may be used to view the drawing surface and the microscope image simultaneously and is helpful for drawing records. The drawing surfaces should be level to avoid distortion. It is desirable to test the drawing arrangement by sketching a hemocytometer ruling, stage micrometer, or other known pattern to make certain that distortion is avoided. This is important when the outlines are made on wax plates or other materials that are to be combined later to give solid reconstruction models (*4,56*).

Photomicrographic records can be made using a light-sensitive plate or film instead of drawing paper and a camera or other dark

enclosure between the microscope ocular and the film. With color
filters and color-sensitive emulsions, considerable flexibility can be
had in photographic records for the emphasis of important details
and the suppression of others. The greater the knowledge of both
microscopy and photography the better will be the records; of the
two a knowledge of microscopy is more important since the usual
principles of photography apply (7,10).

References

TEXTS AND GENERAL REFERENCES WITH GOOD BIBLIOGRAPHIES

1. Beck, C., *The Microscope: Theory and Practice.* 2nd ed., R. and J.
   Beck, Ltd., London, 1938.
2. Carpenter, W. B., *The Microscope and Its Revelations.* 8th ed., revised
   by W. H. Dallinger, Blakiston, Philadelphia, 1901.
3. Chamot, E. M., and C. W. Mason, *Handbook of Chemical Microscopy.*
   2nd ed., Wiley, New York. I (General), 1938. II (Microchemical),
   1939.
4. Gage, S. H., *The Microscope.* 17th ed., Comstock, Ithaca, N. Y., 1941.
5. Mayall, J., Jr., *Cantor Lectures,* "The Microscope; The Modern Micro-
   scope," *J. Soc. Arts,* **34,** 987, 1007, 1031 (1886); **36,** 1149, 1164 (1888).
6. McClung, C. E., *Handbook of Microscopical Technique.* 2nd ed., Hoe-
   ber, New York, 1937. (Forthcoming 3rd ed. will contain sections on
   polarizing and phase microscopy.)
7. *Photomicrography.* 14th ed., Eastman Kodak Co., Rochester, 1944.
8. Richards, O. W., *The Effective Use and Proper Care of the Microscope.*
   American Optical Co., Buffalo, 1949.
9. Richards, O. W., *The Effective Use and Proper Care of the Microtome.*
   American Optical Co., Buffalo, 1949.
10. Shillaber, C. P., *Photomicrography in Theory and Practice.* Wiley, New
    York, 1944.
11. Spitta, E. J., *Microscopy, the Construction, Theory and Use of the Micro-
    scope.* 3rd ed., Murray, London, 1920.
11a. Wredden, J. H., *The Microscope, Its Theory and Applications.* Grune
     & Stratton, New York, 1948.

Journals and Societies

12. American Microscopical Society, *Transactions.* Dr. F. E. Eggleton,
    Secretary and Editor, University of Michigan, Ann Arbor, Mich.
13. Biological Photographic Association, *Journal.* Mr. Lloyd Varden Sec-
    retary, c/o Pavelle Color Incorporated, 553 W. 57 St. New York 19,
    N. Y.
14. Royal Microscopical Society, *Journal.* B. M. A. House, Travistock
    Square, London W C 1, England.

RESOLUTION

15. Magnan, C., P. Chanson, and A. Ertaud, *Compt. rend.*, **220,** 770 (1945) (proton microscope).
16. Merling-Eisenberg, K. B., *J. Quekett Microscop. Club*, **1,** 311, 322 (1937).
17. Lucas, F. F., *J. Franklin Inst.*, **217,** 661 (1934).

INFRARED

18. Clark, W., *Photography by Infrared*. 2nd ed., Wiley, New York, 1946.
18a. Bailly, R., *Science*, **108,** 143 (1948); *Am. Mineral*, **33,** 519 (1948).

ULTRAVIOLET

19. Brumberg, E. M., *Nature*, **152,** 357 (1943); *Compt. rend. acad. sci. U. R. S. S.*, **52,** 499 (1946).
20. Caspersson, T., *J. Roy. Microscop. Soc.*, **60,** 8 (1940).
21. Köhler, A., *Z. wiss. Mikroskop.*, **21,** 129 (1904).
22. Lavin, G. I., *Rev. Sci. Instruments*, **14,** 375 (1943).
23. Trivelli, A. P. H., and R. P. Loveland, *J. Optical Soc. Am.*, **20,** 97 (1930). A. P. H. Trivelli and L. V. Foster, *ibid.*, **21,** 124 (1931).
24. Hámos, L. v., and A. Engström, *Acta Radiol.*, **25,** 325 (1944).

OPTICAL STAINING

25. Royer, G. L., C. Maresh, and A. M. Harding, *J. Biol. Phot. Assoc.*, **13,** 123 (1945).
26. Crossman, G., *Anat. Record*, **99S,** 30 (abstract) (1947); *J. Optical Soc. Am.*, **38,** 417 (1948); *Stain Technol.*, **24,** 61 (1949).

POLARIZATION MICROSCOPY

27a. Drabkin, D. L., *J. Biol. Chem.*, **164,** 703 (1946).
27b. Waugh, D. F., and F. O. Schmitt, *Cold Spring Harbor Symposia Quant. Biol.*, **8,** 233 (1940).
28. Mehl, J. W., *ibid.*, **6,** 218 (1938).
29. Picken, L. E. R., *Biol. Revs. Cambridge Phil. Soc.*, **15,** 133 (1940).
30. Schmidt, W. J., *Die Bausteine des Tierkörpers in polarisiertem Lichte*. F. Cohen, Bonn, 1924.
30a. Jelley, E. E., "Microscopy," in *Physical Methods of Organic Chemistry*. A. Weissberger, ed., 2nd ed., Interscience, New York, 1949, Chap. XV.

FLUORESCENCE MICROSCOPY

31. Richards, O. W., in *Medical Physics*. O. Glasser, ed., Year Book Publishers, Chicago, 1944. (See forthcoming 2nd ed. for more detail.)
31a. Keller, R., and B. V. Pisha, *Clin. Med.*, **54,** 394 (1947).
32. Ellinger, P., *Biol. Revs. Cambridge Phil. Soc.*, **15,** 323 (1940).
33. Haitinger, M., *Fluorescenzmikroskopie*. Akadem. Verlagsgesellschaft, Leipzig, 1938.

34. Harvey, E. N., and A. M. Chase, *Rev. Sci. Instruments*, **13**, 365 (1942) (phosphorescence microscope).

INTERFERENCE MICROSCOPY

35. Tolansky, S., *J. Sci. Instruments*, **22**, 161 (1945); *Phil. Mag.*, **37**, 390 (1946).
36. Grenham, C. G., *Australian J. Sci.*, **9**, 26 (1946).
37. Merton, T., *Proc. Roy. Soc. London*, **A189**, 309 (1947).
38. Williams, W. E., *Applications of Interferometry*. Dutton, New York, 1930.

PHASE MICROSCOPY

39. Bennett, A. H., H. Jupnik, H. Osterberg, and O. W. Richards, *Trans. Am. Microscop. Soc.*, **65**, 99 (1946); A. H. Bennett, *Sci. Monthly*, **63**, 191 (1946).
40. Richards, O. W., *Cold Spring Harbor Symposia Quant. Biol.*, **11**, 208 (1947); *J. Biol. Phot. Assoc.*, **16**, 29 (1947); *Stain Technol.*, **23**, 55 (1948).

MICROMANIPULATION

41. Barber, M. A., *Philippine J. Sci.*, **B9**, 307 (1914); *J. Roy. Microscop. Soc.*, **1915**, 71.
42. Policard, A., *La méthode de la microincineration*. Actualités scientifiques et industrielles, No. 765, Hermann, Paris, 1938.
43. Boyd, G. A., *J. Biol. Phot. Assoc.*, **16**, 65 (1947).

SPECIALIZED MICROSCOPES

44. Lucas, F. F., *J. Franklin Inst.*, **217**, 661 (1934).
45. Graton, L. C., and E. B. Dane, Jr., *J. Optical Soc. Am.*, **27**, 355 (1937).
46. Harvey, E. N., *Science*, **77**, 430 (1933); *J. Franklin Inst.*, **214**, 1 (1932).

ILLUMINATION

47. Dempster, W. T., *J. Optical Soc. Am.*, **34**, 695, 711 (1944).

REFLECTING MICROSCOPES

48. Bouwers, A., *Achievements in Optics*. Elsevier, New York, 1946.
49. Burch, C. R., *Proc. Phys. Soc. London*, **59**, 41, 47 (1947).
50. Johnson, B. K., *Proc. Phys. Soc. London*, **51**, 1034 (1939); **53**, 714 (1941).
51. Snook, H. C., U. S. Pats. 2,153,010 (1939); 2,218,270 (1940).

ACCESSORIES

52. Jelley, E. E., *J. Roy. Microscop. Soc.*, **56**, 101 (1936); **62**, 93 (1942).
53. Alber, H. K., and J. T. Bryant, *Ind. Eng. Chem.*, **12**, 305 (1940).
54. Olson, C., *J. Lab. Clin. Med.*, **27**, 939 (1942).

55. Berkson, J., T. B. Magath, and M. Hurn, *J. Am. Stat. Assoc.*, **30,** 414 (1935).

56. Lison, L., *Les methodes de reconstruction graphique.* Actualités scientifiques et industrielles, No. 553, Hermann, Paris, 1937.

COMMERCIAL SOURCES

57. American Optical Co., Scientific Instrument Division, Buffalo 15, N. Y.
58. Bausch and Lomb Optical Co., Rochester 2, N. Y.
59. R. and J. Beck, Ltd., 69 Mortimer St., London, W 1, England.
60. Cooke, Troughton, and Simms, Kingsway North, York, England. R. Y. Fener Co., 110 Pleasant St., Malden, Mass. (U. S. agents).
61. Corning Glass Works, Corning, N. Y.
62. Eastman Kodak Co., Rochester 4, N. Y.
63. Farrand Optical Co., Bronx Blvd. and East 238th St., New York 66.
64. Gaertner Scientific Corp., 1210 Wrightwood Ave., Chicago 14, Ill.
65. E. Leitz, Inc., 304 Hudson St., New York 13.
66. C. Reichert, Vienna, Austria.
67. W. Watson and Son, 14 Hadley Grove, High Barnett, London, England.
68. Carl Zeiss, Inc., 485 Fifth Ave., New York.

ELECTRON MICROSCOPY

JAMES HILLIER, *RCA Laboratories*

A. FUNDAMENTAL PRINCIPLES OF ELECTRON MICROSCOPY

1. Fundamentals of the Instrument

The electron microscope is an instrument developed by physicists and electronic engineers as a solution to a long-standing problem of the designers of optical microscopes. Over seventy years ago Ernst Abbe pointed out that it was not the lack of perfection in the lenses of the microscope that imposed the ultimate limitation on its resolv-

ing power but rather that the light used, being periodic in nature, obeyed the laws of diffraction. On calculating the diffraction effects for a microscope, Abbe found that it was incapable of resolving details in the object separated by much less than half the wavelength of the illumination used. While there has been much discussion in the literature as to the precise meaning of Abbe's theory, and as to the exact method by which it should be applied, the fact remains that his well known formula $d = 0.61\lambda/(n \sin \alpha)$ provides a reasonably accurate estimate of the resolving power of a microscope system (n is the refractive index of the medium in which the object is placed, α is the angle subtended at a point in the specimen by a radius of the objective, and λ is the wavelength of the light used). The common parameters for an optical microscope using visible light and a good oil immersion objective lead to a resolving power of approximately 0.2 μ (2000 A.). If we standardize on 0.2 mm. as the resolving power of an unaided eye at a viewing distance of 25 cm., it is apparent that the *useful* magnification of such a system is approximately 1000×.

It is immediately obvious that there are many structures found in nature that are beyond the resolving power of such a microscope— the finer structures of cells and bacteria, the entire particle in the case of a virus or a large molecule, have dimensions that lie below 200 mμ (2000 A.). At the same time many of these materials have heterogeneous physical structures that cannot be investigated at all successfully by any of the indirect statistical techniques of biochemistry and biophysics.

Owing to the natural limitations of refractive materials it has has not been possible to make any outstanding increase in the resolving power by increasing the numerical aperture of the objective system. The remaining possibility, that of using a radiation with a shorter wavelength, was exploited very shortly after Abbe's work. Unfortunately, here too the properties of existing refractive materials have prevented the use of radiation beyond the ultraviolet region. With the ultraviolet microscope resolving powers of better than 1000 A. or, in other words, useful magnifications greater than 2000× are not possible. This was indeed a small gain compared to the two orders of magnitude that were really desirable.

The electronic solution of the problem came through a rather involved series of developments by physicists who were primarily interested in the study of the properties of electrons and electron beams. By the year 1924, it was well known that cathode ray beams

produced in discharge tubes or obtained by applying a high positive potential to an electrode near the tungsten filament heated to incandescence consisted of corpuscles of negative electricity known as electrons.   It was also known that those electrons could be deflected by a transverse magnetic or electrostatic field and that the electric fields produced between coaxial cylinders or the magnetic field produced by a current passing through a coil of wire wound coaxially with the electron stream would concentrate the beam into a small point.   While the literature contained some casual remarks concerning the striking similarity between the behavior of the cathode ray streams in such magnetic or electrostatic fields and rays of light traveling through a lens system, the real significance of the similarity was not realized until some years later.

In 1924, however, de Broglie predicted theoretically that a moving electron should have a periodicity associated with it.   This meant that there could be assigned to a moving electron a wavelength having a value given by the formula $\lambda = (150/V)^{1/2}$, where λ is the wavelength in Angstroms (10^{-8} cm.) and V is a measure, in volts, of the velocity of the electron.   Thus, electrons accelerated by 60,000 v. will have a wavelength of 0.05 A., which is only 1/100,000th that of visible light.   This idea was developed theoretically and became the basis of our modern quantum mechanics.   The correctness of de Broglie's theory was demonstrated experimentally by G. P. Thompson and by Davidson and Germer around 1927.   These workers also measured the wavelength of the electron stream and showed that it agreed with de Broglie's calculations.

Almost simultaneously with these developments, H. Busch was working on the action of a rotationally symmetrical magnetic field on an electron beam traveling along the axis of symmetry and in 1926 and 1927 published papers demonstrating that such magnetic or electrostatic fields acted as true lenses, *i.e.*, they were capable of transforming a beam of electrons diverging from a point on the axis of symmetry of the system into a beam converging to or diverging from a second point on the same axis.   He also showed that one could define focal lengths for such lenses and that they could be expressed in a form identical with that already in use for optical lenses.

In pointing out that a rotationally symmetrical electric or magnetic field was a true electronic lens, Busch initiated the science of electron optics.   Very shortly afterward other workers showed theoretically that the analogy between electron optics and light optics was quite

complete and, in fact, stemmed from the basic analogy between opti-
cal and mechanical systems first published by W. R. Hamilton many
years earlier. The fundamental and practical significance of this
theory is that electron-optical counterparts of all optical instruments
are physically possible. Applied to the microscope it leads imme-
diately to the possibility of building an electron microscope. Simi-
larly the theoretical resolving power of such an instrument would be
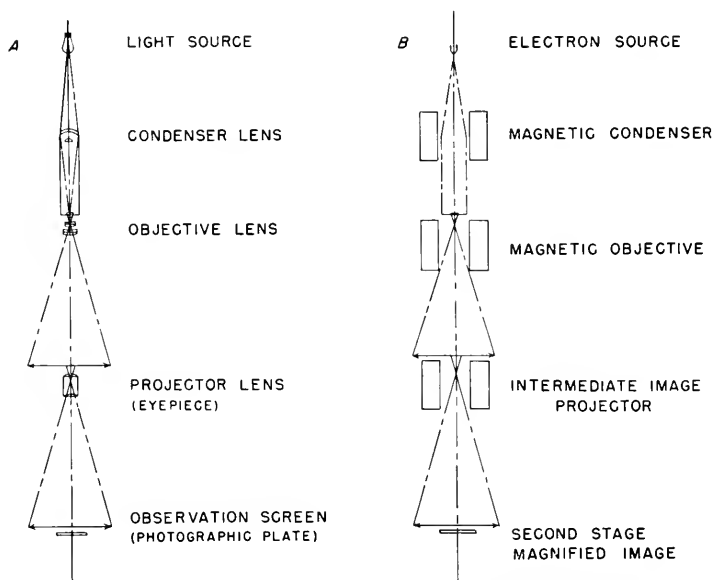
Fig. 1.   Simplified diagram showing analogy between
optical systems of light and electron microscopes.

determined by Abbe's formula and hence, because of the short wave-
length of electrons, would be very high. There were, however, in-
numerable technical difficulties to be overcome before such resolving
power could be realized and even today, when the resolving power of
the electron microscope is as much as 200 times greater than that of a
light microscope, the theoretical possibilities of an electron beam are
by no means fully utilized.

The first attempt to build a high resolving power electron micro-
scope for the examination of materials that were not themselves
electron emitters was made by M. Knoll and E. Ruska in 1932. In
1935, E. Driest and H. O. Müller, working with another instrument

built by E. Ruska, achieved a resolving power somewhat greater than that possible with the light microscope. The development was continued until by 1940 resolving powers of 20 A. had been achieved by a number of workers and electron microscopes were made available commercially. Further improvements in the resolving power were not forthcoming until, in the latter part of 1945, the present writer designed a special objective able to resolve approximately 10 A. (1 mμ).

As would be expected from the preceding, the electron microscope is quite analogous to the light microscope as far as the arrangement of its lens system is concerned (see Fig. 1). However, in its physical structure, it is an entirely different instrument. Figure 2 is a photograph of a modern electron microscope, while Figure 3 is a somewhat simplified schematic cross section of this instrument. The most obvious physical characteristic of an electron microscope is that the entire lens system including the specimen and the photographic plate or film is enclosed in a high vacuum chamber. In view of the fact that it is necessary to align the lenses, to position the specimen, to manipulate the shutter and photographic plate in the vacuum chamber, and at the same time to be able to gain easy access to the inside of that chamber for replacing the specimen or replacing a photographic plate, a very special type of structure is necessary. Furthermore, it is necessary to provide a fast pumping system, which will exhaust the air rapidly from the instrument after changing the specimen or plate and keep it exhausted during its operation. As little as ten years ago, many of the vacuum techniques used in the electron microscope at the present time were not possible, even in the laboratory, but fortunately—due to the simultaneous development of such techniques for many other purposes—they have now become routine and seldom provide any great difficulty.

A complete description of the electron microscope and the details of the construction of its various components cannot be given here. These can be found in any of the general references. There are, however, a number of points in connection with the operation of the electron microscope that should be given in order to clarify the understanding of its use. For instance, the lenses used do not have fixed focal lengths, particularly in the case of those instruments using magnetic lenses. Since the focal length of the magnetic lens is controlled by the strength of the field produced in the lens and that in turn is controlled by the current passing through
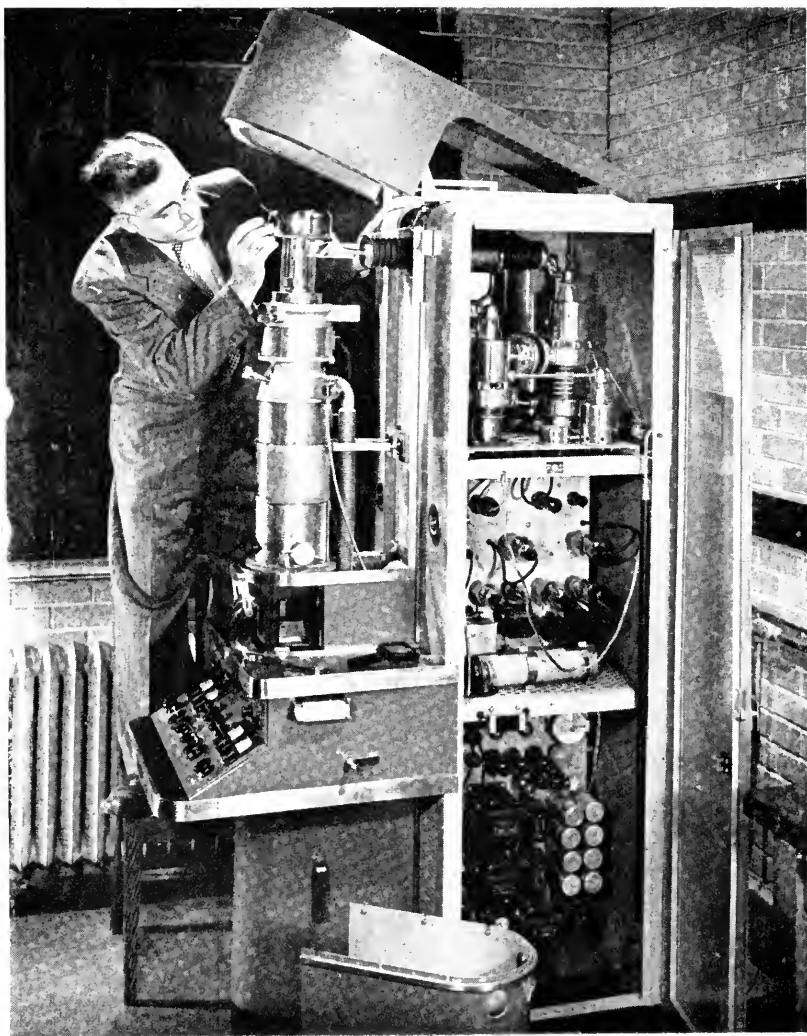
Fig. 2.   Photograph of a modern electron microscope with doors and covers
open, showing actual operating components.

the exciting winding, the operator can vary the focal length of such
a lens over a wide range.   This means that it is not necessary to
move the lenses of an electron microscope to focus the instrument.
Nor is it necessary to replace eye lenses or objectives if different

magnifications are desired.   Similarly, the substage condenser, unlike
its light-optical counterpart, is not usually provided with an iris
diaphragm.   Nevertheless, the angular aperture of the illumination
can be controlled over a wide range of values by varying its focal
length.   The magnification of the objective of an electron micro-
scope is a fixed quantity determined by the positions of the specimen,
objective lens, and projection lens.   On the other hand, the mag-
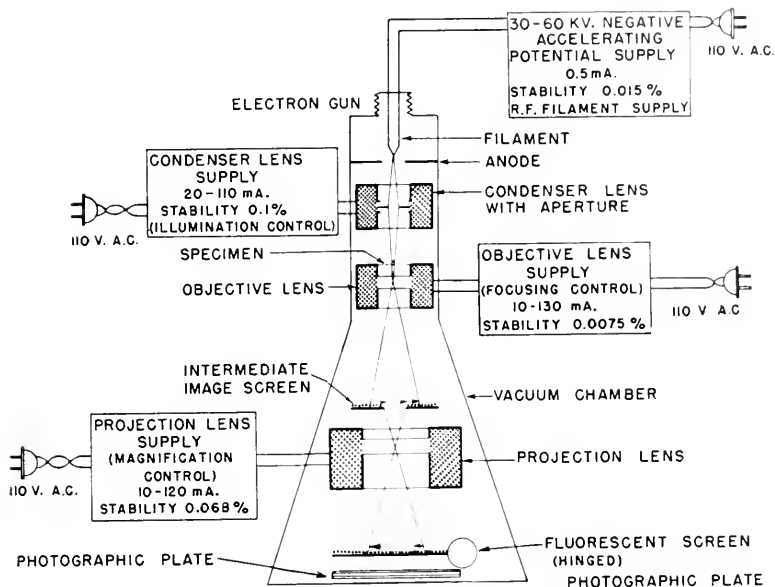
Fig. 3.   Simplified cross section of magnetic type of electron microscope.

nification of the projection lens can be varied over a wide range by
adjusting the current in the exciting coil.   The available range may
be as much as 4 to 1.   If a greater range of magnification is desired,
they can be achieved by replacing the polepieces of the projection
lens or through the provision of a third lens, which—when varied
appropriately in conjunction with the normal projection lens—can
provide variations in magnification over a range of 500 to 1.

In the light microscope all of the useful magnification is usually
obtained directly in the instrument.   In the electron microscope this
is seldom done for a number of purely technical reasons.   The first
of these is that at the highest magnifications the field of view is

extremely small unless the instrument is provided with a very large
final viewing screen and photographic chamber. Since the intensity
of the image varies inversely as the square of the magnification, it is
exceedingly low at very high magnifications for illumination inten-
sities the specimen can tolerate. This results in poor visibility for
visual observation and extremely long exposures for photographic
work. In the case of magnetic instruments, the latter requires the
provision of power supplies that remain sufficiently constant over
such long periods. For these reasons and other less important ones,
the magnification obtained electronically is usually kept at the
lowest value consistent with the retention in the final micrograph of
the resolving power provided by the objective lens. Practical in-
strumental magnifications in electron microscopes seldom exceed
$30,000\times$ and may run as low as $1000\times$. Since the finest details of
these images cannot be resolved with the unaided eye, it is customary
to magnify the visible image optically for focusing the instrument.
The photographic recording of the image without loss of details is
accomplished by the use of a fine grain photographic emulsion.
From the known resolving power of the photographic emulsion and
the resolving power desired in the material under investigation, it is
a simple matter to calculate the *minimum* practical instrumental
magnification.

Another important difference between light and electron micro-
scopes is in their depths of field. As can be seen from Abbe's formula,
in the light microscope the angular aperture of the objective is made
as large as possible in order to achieve the highest resolving power.
However, since the depth of field of a microscope is determined by
the expression $D = 2d/\tan \alpha$ (D is the depth of field, d the desired
resolving power, and α the angular aperture of the objective) it can
be seen that the depth of field of the light microscope where α is
nearly $\pi/2$ radians is extremely shallow. In fact, the depth of field
for high power light microscope objectives is usually less than the
resolving power of the instrument. This means that all the prep-
arations under examination must be very accurately in one plane if
the image is to be in focus at all points. Conversely information
regarding a "thick" specimen can be obtained with the light micro-
scope by manipulating the fine focus control during visual observa-
tion. Such information cannot be recorded in a single photographic
record, a fact that explains the conclusion that a photomicrograph

is not as satisfactory as direct observation through a light micro-
scope. In the electron microscope, however, the angular aperture
of the objective is usually very small, less than 0.01 radian. Thus,
in the electron microscope the depth of field is correspondingly large
and may be several microns for work requiring moderately good re-
solving power. This is actually an important advantage for the
electron microscope since, for any specimen suitable for examination
in the electron microscope, the entire field of view is in focus even
though it may have considerable depth. As a result of this and
because of other technical reasons electron micrographs, that is,
photographically recorded electron microscope images, usually con-
tain more information than the images made visible by means of a
fluorescent screen. This also makes it possible to obtain stereoscopic
images providing three-dimensional pictures of the specimen at the
highest magnifications. Unfortunately, because all techniques in
light microscopy have been developed with these limitations in mind,
it will be necessary to develop new techniques before the full useful-
ness of the electron microscope in this extra dimension can be
realized. Nevertheless, it is a point that the present or prospective
users of an electron microscope should keep in mind, because very
often, even with available techniques, it is possible to obtain micro-
graphs with the electron microscope that show a manifold improve-
ment over those obtained with the light microscope, even though the
extremely high resolving power of the electron microscope may be
only partly utilized (Fig. 4).

One of the most important differences in the results provided
by the light and electron microscopes is found in the structures of
the images themselves. In the light microscope the image exists
by virtue of the fact that different points of the specimen are different
in a number of their physical properties—such as refractive index,
selective absorption, and scattering power for light—and is the result
of a superposition of the effects of these differences. In the electron
microscope, on the other hand, the situation is much simpler. The
image is monochromatic and consists only of intensity variations
resulting from differences in the electronic scattering power of differ-
ent points of the specimen. While this simple relationship is seriously
disturbed by the lens aberrations and by the diffraction of electrons
by crystalline particles in the specimen, it leads nevertheless to a
more direct interpretation of the image.

2. Fundamentals of Application of the Electron Microscope

In the preceding section the fundamental principles involved in the design and operation of the electron microscope have been set up in the most elementary form possible.   To the scientist, whether he be biologist, chemist, or metallurgist, who wishes to apply the electron microscope in his own particular research, that description is of academic interest only.   Actually, he needs comprehend only one fact, namely, that the electron microscope is a device for producing a visual image of the microscopic structure of the material he is studying.   In fact, the use of electrons distinguishes this instrument from other microscopes only insofar as it permits the observation of finer structures in the same way that light microscopy is distinguished from ordinary vision by the dimensions of the structure examined by that method.   Thus, in its true sense, electron microscopy is merely a term defining the simple use of our ordinary sense of vision in research involving a very fine scale of dimensions.

To state that our sense of vision is one of our most important assets in research would seem platitudinous.   At the same time our sense of vision has been so much a part of our environment that we are sometimes inclined to forget that the eye–brain combination is an almost fantastic device capable of measuring and recording in the space of one-tenth of a second the intensity, wavelength, and position of over a million points of light.   By way of comparison consider the indirect scientific method of investigating, say, an everyday scene. Actually such an analysis is carried on electronically in a television camera, where the intensity of each point of the scene is measured and converted to an electrical oscillation of a corresponding amplitude.   The scientific data in this case would consist of a long list of several million numbers representing the amplitude of that electrical signal at regular intervals, say, every ten-millionth of a second.   The manual conversion of such data into a form that would reveal the desired information regarding the original scene would, of course, involve a prohibitive amount of labor.   Thus, the data are valueless unless electronic means are used for reconverting and presenting them in the form of a visual image, which is then comprehended immediately.

The data obtained by a research worker using indirect methods to investigate a heterogeneous system on a microscopic scale are quite analogous to the tabulated data representing the television image.

In general, however, there has been no device available to convert
that data into a form intelligible to the human senses. Instead, the
research worker has had to content himself with making small num-
bers of measurements on systems that he considers to contain large
numbers of only a few different types of identical particles. Such
methods have been ineffectual as soon as those particles become organ-
ized in complex heterogeneous structures. It is in this range that
the electron microscope provides much of the required information in
the form of a visual image.

We can also use the television analogy to illustrate the funda-
mental research problem of electron microscopy. On the television
screen we see images representing scenes with which we have some
familiarity. Within certain limitations they are presented quite
faithfully. If this is not so, the observer's familiarity with everyday
scenes enables him to realize the change immediately and at least
mentally compensate for the deficiencies of the presentation. With
the electron microscope, however, the situation is not the same. The
images presented are of things initially totally unfamiliar to the
research worker. Moreover, he has no experience or information
by which he may judge their accuracy. To obtain that experience
he must then carry on by varying his technique of preparation,
performing other experiments, observing the results, and accumulat-
ing the necessary experience as he progresses. In fact, most of the
preparatory work involved in applying the electron microscope to a
particular problem is concerned with the development of a method
of preparing the specimen in such a way that the images produced
can be considered reliable. Once the research worker has convinced
himself that his images are true representations of the specimen, he
is ready to use his instrument as a means of visually observing the
results of experiments that concern his original research, and electron
microscopy *per se* plays no further role in his work.

Two of the ideas implicit in the above discussion, which might be
emphasized further, concern the precise definition of the nature
of the information obtained with the electron microscope. It should
be realized that visual information obtained from a monochromatic
image involves only the geometry of an object. As mentioned above,
the intensity differences in an electron microscope image are the result
of differences in scattering power of the individual points of the speci-
men, so that there is some correlation between the intensity of an
image point and the physical properties of the corresponding speci-

men point.   However, this additional information can be used only
with caution since there are many reasons for its being unreliable.
Obviously, the information provided in an electron microscope image
is *not* statistical in nature.   This is in contrast to the information
provided by practically any other kind of scientific instrument.
By its very nature, the electron microscope is an instrument for ex-
amining heterogeneous samples.   While it can examine homogeneous
materials, it seldom can provide more accurate information than can
be obtained by other techniques.

B. FIELDS OF APPLICATION OF ELECTRON MICROSCOPY

1. Introduction

Electron microscopy has made it possible to observe the structure
of solids in the range of dimensions from 1 millimicron to 1 micron.
It is unfortunate that such a direct and simple statement, accurate
though it may be, gives no suggestion of the true magnitude of the
role that electron microscopy can play in any science that involves in
some way the structure of solid materials.   In fact, to this writer, it
appears that there are few problems in chemistry, physics, metallurgy,
mineralogy, geology, biology, medicine, etc. in which the electron
microscope cannot be used.   At the same time it is obvious that the
electron microscope does not yet play such an important and wide-
spread role in general science.   The reasons for this are also quite
clear.   The first of these is the lack of information among the workers
who might make use of an electron microscope as to its capabilities
and limitations.   The second is the fact that early difficulties en-
countered with the instrument and with some of the associated
techniques have caused the more conservative research workers to
wait until those difficulties have been overcome.   On the practical
side, one of the most important controlling factors in the general use
of the electron microscope has been, as already mentioned, the lack
of suitable techniques of specimen preparation.

At the present time, a discussion of the general problem of preparing
electron microscope specimens is probably the best means of providing
the reader with an evaluation of the present capabilities of the
instrument.

2. General Types of Preparations Suitable for Electron Microscopy

Specimens suitable for examination with the electron microscope must satisfy three conditions:

(a) The specimen must be sufficiently thin. If a high resolving power electron microscope image is to be obtained, it is necessary that the electrons penetrate the specimen without appreciable loss in energy. This means that the thickness of the specimen, plus its support, must be less than some definite value. It is difficult to give an exact value for this permissible thickness since it is determined by the nature of the problem as well as the properties of the specimen and the velocity of the electron beam. For instance, the problem may be the determination of the shapes and sizes of metallic particles. Since silhouettes only are of interest, the particles may have dimensions from several microns down. If, on the other hand, the problem is the internal structure of such particles, they must be less than a hundred millimicrons in thickness if they are to be observed by transmitted electrons. In biological material, where the density is seldom very great, layers, including the supporting membrane, of less than a quarter of a micron are usually quite satisfactory.

(b) The specimen must provide sufficient contrast. Since the electron microscope image exists only by virtue of the differences in the scattering of electrons by different points of the specimen, it is necessary that the structures to be observed produce such differences in the scattering. For instance, isolated bacteria mounted on a thin collodion membrane will give an image of ideal contrast as far as the contours of the organisms are concerned, while bacteria mounted on a collodion membrane but covered by a uniform organic film equal in thickness, say, to the height of the organism will provide an image in which the bacteria are hardly distinguishable, if at all. Similarly, in the first case, structures of nearly equal densities within the protoplasm of the individual cells will be indistinguishable, unless it is possible to enhance the contrast by some technique of specimen preparation or by some instrumental development (Fig. 5).

(c) The specimens must be prepared in such a way that they are unaffected either by the vacuum or by the electron bombardment. Since the specimen is placed in a vacuum, it is obvious that it must contain no volatile liquid. This eliminates the possibility of examining undried specimens of any type. In passing, it is to be noted that
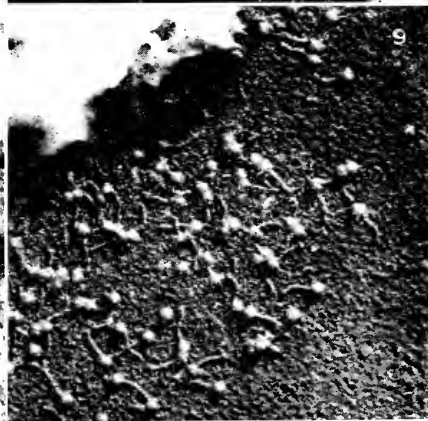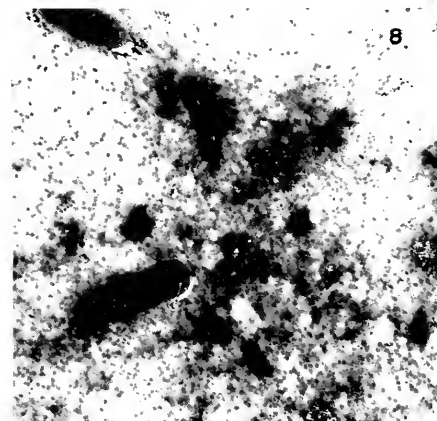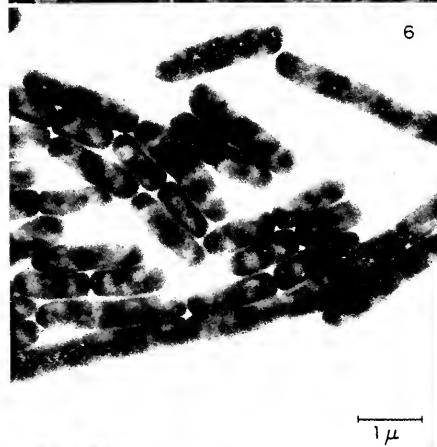
while this limitation is imposed by the necessity of placing the speci-
men in a vacuum, if this were not the case, it would be imposed by
the necessity of attaining sufficient contrast. Most wet biological
preparations would produce little, if any, contrast when examined
in the electron microscope. The second part of this condition,
namely, the changing of the structure by electron bombardment,
might also be considered as part of condition *a*, but has been inserted
here because of the nature of the changes observed. As stated above,
the reason for making specimens sufficiently thin is to prevent the
electrons transmitted by it from losing energy, which in that case
introduces aberrations in the image. However, absorption of energy
from the electron beam results in a rise in the temperature of the
specimen and, hence, can result in the changing of the structure by
evaporation of the somewhat more volatile materials or by chemical
or crystallographic changes.

Fig. 4.   Electron micrograph of portion of underside of a leaf showing single
          stomata. Specimen was prepared by a silica replica technique and
          because of the very rough leaf surface shows some artifacts as a result of
          folding of the replica.   In spite of this great roughness, all parts of the
          replica are in equally good focus (950×).

Fig. 5a. Electron micrograph of *Bacterium mycoides* grown on nitrogen-free
          medium to make cells transparent. Made with standard objective lens
          operated without a diaphragm (5600×).

Fig. 5b. Electron micrograph of same field of view as Figure 5a but made with
          double objective and with an extremely small diaphragm (5600×).
          (Specimen prepared by Dr. G. Knaysi, Cornell University.)

Fig. 6.   Electron micrograph of normal cells of *Escherichia coli*, strain B,
          grown on supporting membrane by technique described in text and
          examined with electron microscope using double objective with a small
          diaphragm (5000×).

Fig. 7.   Small colony of *Escherichia coli* grown as described, examined one-half
          hour after infection with dilute suspension of T_2 bacteriophage particles.
          Titer of bacteriophage inoculum was so adjusted that each colony
          would be infected by one bacteriophage particle. Lysis of a single cell
          is shown in the micrograph. Several bacteriophage particles are visible
          in remains of the cell but surrounding cells appear normal (9500×).

Fig. 8.   Preparation of *Escherichia coli* similar to that of Figure 6, but ex-
          amined eighteen to twenty hours after infection with T_2 bacteriophage.
          Only a few nonlysed cells can be seen, but field is covered with very
          large number of bacteriophage particles and lysed cell debris (7200×).

Fig. 9.   Electron micrograph of bacteriophage in a *Streptococcus lactis* prepara-
          tion shadow cast with gold to give a shadow four times longer than
          height of the organisms (30,000×).   (Courtesy *Iowa State College*.)

At first sight it would appear that these conditions place quite stringent restrictions on the range of objects that can be examined with the electron microscope. Actually, that has not been the case and while many of the techniques developed have required considerable imagination and skill, it has been possible nevertheless to satisfy these conditions in a surprisingly large number of cases. After several years of seeing the "impossible" problems of specimen preparation solved by simple means, this writer is convinced that these conditions can be fulfilled for most of the materials that may be encountered. The conditions discussed above limit the use of the electron microscope to the examination of solid specimens. At the same time it now appears that few, if any, solids cannot be studied with that instrument. Obviously the field of application is an enormous one and a thorough discussion is beyond the scope of this article. It is hoped that the general discussion in the above and in what follows will allow the reader to judge for himself the utility of the electron microscope in any particular problem.

Solids to be examined by means of the electron microscope can be divided into two broad groups—the particulate and the compact solids. This subdivision is made almost entirely on the basis of the method of preparation of the specimen, since the information sought in either group is the same.

3. Examination of Particulate Specimens

Many specimens to be examined with the electron microscope are naturally particulate; others are made particulate by fragmentation of a compact structure.

There are many methods of preparing particulate specimens, which vary according to the material. All of them involve obtaining a suitable dispersion on or in a thin supporting membrane or on some other support without changing the particles in any way. For smokes, soils, pigments, colloids, catalysts, macerated fibrous materials, etc., we find that condensation from a vapor stage, thermal or electrostatic precipitation from an aerosol on a plastic membrane, milling into a membrane-producing solution, centrifugation or electrophoretic deposition from liquid suspension, or the simple evaporation of a liquid suspension, are the main methods used. For large organic molecules and viruses, centrifugation or evaporation from a purified suspension are useful methods. For bacteria or other single cells,

evaporation from aqueous or saline suspensions has commonly been used, though more recently techniques of growing these directly on the supporting membrane have been developed. Since the latter technique seems to be much less susceptible to the introduction of artifacts, it will probably soon replace the older technique.

All the above methods of preparation lead to micrographs that provide two dimensional representations of the specimen. In the majority of cases this is sufficient, since the three dimensional representation can be inferred from the appearance of the particles and from a consideration of the method of preparation. In some cases, however, the problem requires precise three dimensional information that cannot be deduced in this way. It is then necessary to make a direct determination of the three dimensional shape of the particles. Two methods are now available.

The first is direct stereoscopic imaging of the specimen. Owing to the great depth of field of the electron microscope it is possible to obtain high magnification stereomicrographs by the simple method of making two micrographs of a given field of view, the specimen having been tilted, between exposures, about an axis that lies in its plane. Under ideal conditions this method is capable of resolving approximately 20 mμ in depth.

The second is the recently developed shadow-casting technique (17), which is more suitable for smaller particles. This involves placing the prepared specimen, prior to examination, in a vacuum chamber where a light layer of heavy metal is condensed on its surface. The heavy metal is evaporated in the vacuum chamber from a small source, usually a heated filament, basket, or foil boat, so placed that the vapor strikes the specimen obliquely. The condensed layer of metal has then a distribution on the specimen exactly analogous to that which would obtain for the illumination of an enlarged model of the specimen produced by a similarly placed small light source. However, in a positive electron micrograph—due to its greater scattering power—the metal appears darker in those regions where it is more heavily deposited. For this reason, such micrographs are usually printed as negatives whence they take on the appearance of an obliquely illuminated surface, permitting direct observation of the relief present in the specimen. The method is particularly valuable for showing the existence and shape of small particles that by virtue of their size and density are normally very nearly indistinguishable from the supporting membrane, for resolving

the fundamental ambiguity that always exists in the electron micrographs between the density and thickness of a specific specimen point, and for showing surface characteristics of larger objects where the thickness of the underlying material has reduced the contrast or that underlying material has an internal structure that masks the structure of the surface. See Figure 9.

The last point indicates a major limitation of the shadow-casting technique, particularly for biological work. Information obtained through the method of shadow-casting is confined to the infinitely thin outer surface of the specimen and usually completely submerges the image of any existing internal structure. It is only in the very specially prepared biological specimens that the surface carries significant information. For this reason, the writer considers the present practice in some laboratories of shadow-casting all specimens a dangerous procedure leading both to erroneous interpretations and to an apathy toward the basic problems of structure determination.

4. Examination of Compact Solids

Coincident with development of the electron microscope as a commercial instrument, its inability to examine surfaces was presented as one of its major limitations. This arose from the fact that sufficiently thin sections could not be cut of most materials and that electronic images by reflection from a surface could not be made successfully owing to the wide range of velocities present in the reflected electron beam. However, this apparent difficulty was overcome almost immediately by the development of what is known as the replica technique by H. Mahl in Germany and V. K. Zworykin and E. G. Ramberg in the United States. Basically, this technique involves the production of a thin plastic replica of the surface to be examined. The replica is made thin enough to be suitable for examination in the electron microscope. The geometrical details of the surface structure are made visible by the fact that at different points the replica has different thicknesses and, hence, a different scattering power for the electrons. At first it seemed unlikely that such a replica could reproduce the surface structure on a sufficiently fine scale. At present, however, it has been well established that the chief limitation in the technique is the molecular size of the material used in making the replicas. Since this is usually considerably less than 10 mμ, the method is quite satisfactory for resolving powers of

this order.   While the technique was developed initially for metallo-
graphic problems and, in fact, the early techniques could not be ap-
plied elsewhere, the general technique has been continuously varied
and improved so that at the present time it is possible to make replicas
of all solid surfaces except possibly a few extremely delicate and semi-
liquid ones.   The study of the external structure of such compact
solids as metals, glasses, ceramics, plastics, paint surfaces, wood, parts
of plants, flowers, seeds, leaves, many parts of insects, colonies of bac-
teria, teeth, hair, skin, blood cells, etc. is now possible.   The replica
technique also makes it possible to examine the internal structure
of many of the same objects.   This is done by making replicas of
internal surfaces exposed by various methods.   Generally speaking,
however, this variation is not applicable to delicate biological struc-
tures; it can be applied to the structure of bones, teeth, etc.

Stereoscopic and shadow-casting methods are very necessary ad-
juncts to replica work since here the three dimensional structure can
seldom be inferred from a single micrograph without a very accurate
knowledge of the nature of the replica and without tedious point-
by-point measurements of the density in the micrograph.   Because
all the information provided by a replica is contained in a surface
and moreover in only one of the two surfaces that determine its
point-to-point thickness and hence its appearance by simple trans-
mission imaging, the shadow-casting technique for demonstrating
the relief is ideal.   It should be mentioned that direct visual inter-
pretations of replicas made without these aids are almost invariably
erroneous.   This is a fundamental limitation due to the normal
tendency of the mind to interpret light and shade in terms of illumi-
nated objects.

To study internal structure of the more delicate biological mate-
rials, it is necessary to use sectioning techniques.   While this has been
initiated, it remains one of the potential fields of application of the
electron microscope in which success has been limited.   The basic
difficulty at present lies in the fact that nearly all the histological
techniques developed for the light microscope have obvious defects
apparent as soon as the resolving power of the light microscope is
only slightly surpassed.   Added to this is the technical difficulty of
cutting sections only a fraction of a micron in thickness without intro-
ducing serious distortion in the tissue.

Very recently, D. C. Pease and R. F. Baker, by the combined use
of very small blocks, double imbedding (hardened collodion followed

by high melting point paraffin), and a simple modification of a stand-
ard rotary microtome, succeeded in producing, routinely, sections
only 0.1 μ thick and quite suitable for electron microscope studies (*18*).

C. ELECTRON MICROSCOPES, AUXILIARY APPARATUS, AND TECHNIQUES

1. Number of Electron Microscopes

It has already been indicated that nearly any laboratory interested
in the heterogeneous microscopic structure of solid materials, regard-
less of their nature, can make profitable use of an electron micro-
scope. While it would appear that practically every scientific labora-
tory that in some way was interested in chemistry, physics, min-
eralogy, metallurgy, geology, many types of engineering, biology,
or medicine would have a need for an electron microscope, it is of
interest to note that the total number in the laboratories of the world
probably does not exceed 400, a number that is being increased at the
rate of roughly one per week.

In addition to those already given, a very important factor limiting
the rate of growth of the use of the electron microscope is the shortage
of trained personnel. While this shortage has existed for several
years in all fields of science, it has been particularly acute in the case
of electron microscopists. Most laboratories that have recognized
their need for an electron microscope have also realized that it will
be applicable to a number of the different problems being investigated
in their laboratories. At the same time, they have been aware that
it would be inefficient and uneconomical—not to mention unsuccess-
ful—for all the individuals working on these problems to interrupt
work to learn the techniques of electron microscopy. It has been
shown to be much more satisfactory to set up the electron microscope
in a separate service group with its own trained personnel. This
results in a minimum requirement of at least one trained electron
microscopist for each instrument. Up to the present time, the
necessary training has been available only through actual experience
with the instrument in a laboratory using it. Since the adequate
training of a microscopist in this way usually takes the better part of
two years, the instruments have been delivered at a much higher
rate than personnel to operate them have become available. This
situation should be remedied in the very near future, since a number

of universities are in the process of arranging graduate courses in
which formal training in electron microscopy will be given.

Another item of importance that often delays the application of
the electron microscope to a problem is its high cost. Electron
microscopes available at the present time vary in cost between
$7000 and $30,000. To introduce an item of this magnitude in
the budget of any scientific laboratory requires a very considerable
justification. In some cases it is relatively easy to find such justi-
fication. For instance, in industrial chemical laboratories there are
many cases on record in which the use of an electron microscope over
a relatively short period has resulted in annual savings amounting
to many times the original cost of the instrument. In universities,
however, a somewhat different situation has existed. In practically
any school there are a number of departments that have problems
to which the electron microscope can be applied. While it is only
in the rare cases that an individual department is in the financial
position to purchase an electron microscope, it is usually possible
for the university as a whole to purchase an instrument, the expenses
being divided appropriately among the departments that make
use of it. Unfortunately, the co-ordination of research among various
departments of a university is not always such that this possibility
is recognized.

Another unfortunate situation that sometimes arises in finding
justification for the purchase of an electron microscope is that of
the laboratory which feels it has a need for an instrument but which
bases its decision not to buy one on a discouraging experience of a
similar laboratory at an earlier time, failing to realize that electron
microscopy is advancing at a very rapid rate and that most of the
major difficulties encountered by the early workers have since been
overcome.

2. Electron Microscopes

At the present time a number of different electron microscopes are
available throughout the world. In Europe, Philips in Holland,
Metropolitan Vickers in England, and a number of other manu-
facturers have designed and either intend to produce or are already
producing electron microscopes. All these instruments have been
designed since the end of the war and relatively few of them are yet
in scientific laboratories. In America, only, has there been contin-
uous production of electron microscopes since before the war. These

instruments have been made by the Radio Corporation of America (RCA) since 1940 and have been improved continuously.

These instruments can be divided roughly into two classes; first, the research instrument, which has rather high resolving power ranging, for various makes, from 10 to 2 mμ, and the smaller instrument, which is simpler to operate but more suited to routine test work, having a resolving power of the order of 10 mμ.

It is not possible at the present time to discuss the relative merits of the different designs. Only partially complete information is available in this country concerning the European designs, which have many novel features that have not yet been tested by widespread use in laboratories outside those of the designers. From the published micrographs produced by the different instruments, it is quite apparent that the most recent RCA instrument has the highest resolving power, though the instrument designed by le Poole as the basis of that to be produced by Philips compares quite favorably with it. At present the RCA electron microscope (Universal Model) appears to have other advantages in the basic nature of its design, which make it considerably easier to adjust and operate, and in the advanced design of the power supplies.

3. Electron Microscope Laboratory

At present, and probably for some time in the future, the electron microscope will be used as a research tool by research laboratories investigating a variety of problems. It has been the experience of such laboratories that the best use can be made of the electron microscope if it is set up as a separate service laboratory operated by a small group of experienced electron microscopists who at the same time have interests in common with those of the other sections of the laboratory. As has already been pointed out, it has been shown to be unsatisfactory to have an electron microscope in a laboratory available for use by the individuals concerned with the actual research problems. In the cases in which such a system has been tried it has been found that none of the individuals using the instruments obtained sufficient experience with it or understanding of it to obtain the full benefit of its use. Furthermore, the maintenance of the instrument itself became a haphazard matter, which generally resulted in its slow degeneration.

In the following will be described what appears to be a suitable, if not ideal, arrangement of an electron microscope service laboratory.

Physical Arrangement. Auxiliary Equipment. A research type electron microscope is a rather bulky piece of equipment standing 7 feet high and weighing several hundred pounds. There are, in addition, as separate units, a mechanical pump and usually a line voltage regulator transformer or some other auxiliary piece of the power supply equipment. The instrument is somewhat susceptible to vibration, though not excessively so. This means that it should be located in a section of the laboratory that is reasonably free from vibration. It is not usually necessary to provide a special vibration-free mounting, though this has occasionally been the case in difficult circumstances where heavy machinery has been installed in nearby rooms or where peculiar geological formations have transmitted vibrations from nearby roads or railroad tracks. The mechanical pump should be mounted as close as possible to the instrument. Since these pumps are usually quite noisy, it is highly desirable to enclose them in a soundproof box or to mount them in an adjacent room or below the floor. It is of course necessary to prevent any vibrations produced by the pumps from affecting the microscope. Usually flexible rubber mountings for the pump are sufficient.

It is essential that electron microscope laboratories include a conveniently located dark room. The most satisfactory arrangement seems to be to have the dark room adjacent to the microscope room and communicating with it through a maze or light lock. The dark room equipment should include a high quality enlarger, a print box, an X-ray type negative viewer, print dryers, preferably of the heater type, and the usual sink with suitable developing trays. The room itself should be large enough to permit two or three people to work in it simultaneously.

The electron microscope room need not be too large, since it is essential that it, too, be arranged so that it can be completely darkened. Generally speaking, the last requirement makes it undesirable to place any other equipment in the same room with the electron microscope. It is essential that the electron microscope room be provided with good ventilation, particularly in those regions of the country that experience hot, humid weather. Most electron microscopes dissipate somewhat more than a kilowatt of power when in operation, which is usually sufficient to make a small room very uncomfortable unless well ventilated. The remainder of an electron microscope laboratory should consist of at least one relatively large, well lighted room. In this would be placed a bench for specimen

mounting and that auxiliary equipment peculiar to an electron micro-
scope laboratory.   The last includes a good quality light microscope,
perhaps supplemented by a microscope of intermediate quality for
rough work.   A comparator should be conveniently available if the
work includes any amount of particle size measurement or electron
diffraction studies.   Because of the recent development and wide-
spread use of the shadow-casting technique an electron microscope
laboratory is not complete without vacuum evaporating equipment.
The electron microscope laboratory should also include ample facilities
for filing negatives and prints, since these generally accumulate with
surprising rapidity.

Personnel.   The personnel of the electron microscope labora-
tory can vary to a certain extent according to the organization of
which it is a part.   A minimum requirement would appear to be a
trained electron microscopist and a technician or a well trained as-
sistant.   The microscopist should be responsible for the mainte-
nance of the instrument and for the quality of the work produced by
the group.   The most successful electron microscopists at the pres-
ent time have been individuals who have had education or experi-
ence equivalent to a doctorate degree or better in physics, stressing
optics or light microscopy.   Such individuals will find some advan-
tage in having had some experience in electronics, but it is by no
means essential.   It is extremely important that the electron micros-
copist have a thorough understanding of the general problems of the
laboratory in which he works.   He is then able to discuss intelligently
the results of the electron microscope investigations and its relation-
ship with the research problem of the individuals providing the
specimens.   It is obvious that the selection of the appropriate individ-
ual for this position is a major step toward the successful utilization
of the electron microscope by any laboratory.   The assistant or
technician, on the other hand, need not have as much experience or
formal training.   It is highly desirable, however, that that individual
be able to learn to make high quality electron micrographs of the
specimens provided and also to carry on the photographic developing
and enlarging that is so much a part of electron microscopic work.

Operation.   In operation the electron microscope laboratory
will be provided with problems from other research groups in the
laboratory or group of laboratories with which it is connected.
It will be the responsibility of the electron microscopist to comprehend
the type of information required for the problem and to use or develop

techniques for specimen preparations that will permit him to obtain that information from the electron microscope images. Finally, he will assist in the interpretation of the images obtained, his main obligation in this connection being the prevention of their being misinterpreted.

Expense. It is obvious from the above that the efficient use of an electron microscope involves considerable expense beyond the initial outlay. While the actual amount of money spent by the different electron microscope groups already in existence varies greatly, it appears to this author that prospective electron microscope users can expect the annual cost of such an organization to be at least equal to the initial cost of the instrument. It should be pointed out that this estimate includes all hidden, as well as apparent, costs. Actually, the total cost is not always the significant one, since in some organizations much of the money spent for the maintenance of an electron microscope laboratory would be spent in any case. In these cases, the major expenditure is that involved in the actual purchase of the electron microscope and the necessary auxiliary equipment.

4. Skill Involved in Operation of the Instrument

In the organization suggested above, the research worker who has need of an electron microscope does not have to be familiar with the technical details of the operation of the instrument, though obviously he should make himself familiar with the fundamental principles involved and with the significance of the images obtained. If, however, this ideal situation is not possible and the research worker finds it necessary to operate the electron microscope himself, it is desirable that he first devote a considerable amount of time familiarizing himself with the general manipulation of the instrument.

Learning to operate an electron microscope is in many ways analogous to learning to drive an automobile. The simple manipulations necessary to make the device function are extremely easy to learn in either case, particularly if the individual has some fundamental knowledge of the principles involved. On the other hand, the new automobile driver is not considered to be a good one until he has driven many thousands of miles on all types of roads and has encountered many traffic problems. Similarly an electron microscope operator does not become proficient at producing high quality micrographs or at recognizing defects and their causes until he has made,

and studied objectively, many hundreds of exposures. Too much emphasis cannot be placed on the necessity of taking large numbers of exposures and correlating the appearance of the image at very high magnifications with the mode of operation of the instrument. It is the opinion of the writer that the only really essential prerequisites to the training of an electron microscopist is that he have sufficient natural aptitude for the handling of scientific equipment plus the basic knowledge of the optics of any microscope. Training in physics and electronics will be helpful but only in that such knowledge will permit him to discover more rapidly the cause of the improper functioning of any part of the instrument.

5. Skill Involved in Preparation of the Specimen

If a worker considers the possibility of utilizing an electron microscope in the investigation of the structure of a material he is studying, it is to be presumed that that worker is quite familiar with many properties of that material. Because of this it is very difficult to differentiate between the skill required by the research worker in the normal execution of his investigation and that required to prepare that specimen in a form suitable for electron microscopy.

There are, however, a few technical aspects of the preparation of electron microscope specimens that may require special skill. For instance, such a preparation almost always involves the handling of very thin film, which must be produced and mounted on wire screens or apertured discs. The specimen must be manipulated in such a way that ultimately it is present on this film and must be thin and free of extraneous materials. Fortunately, many satisfactory techniques for handling these materials have been worked out and are described in the literature. None of them is as difficult as might be expected. Beyond this the research worker is simply handling, in perhaps a somewhat different manner, material with which he is familiar. One might consider, for example, the case of a bacteriologist. He is quite familiar with the methods of making films and of fixing and staining the organism for light microscope examination. Actually, he can carry out identical manipulations for the electron microscope preparation with the one exception that he covers the microscope slide with a thin plastic film that is later to become his supporting membrane. After he has completed his manipulation, he floats this membrane off the glass slide and onto the surface of water and picks it up on a wire screen. If, as he would most likely

find, the treatment he has given the organism, while satisfactory for the light microscope, has introduced many fine scale artifacts that can be detected in the electron microscope, he may then develop new techniques such as that of growing the organisms directly on the plastic membrane or of removing the organisms intact from the surface of a culture. Having developed the techniques so that he is satisfied with the authenticity of the resulting images, he returns to his original research problem.

6. Analysis of Electron Microscope Results

Research results obtained from electron micrographs take two forms. The most obvious and direct of these are measurements of size and shape of the particles and structures recorded in the image. Of greater importance are the qualitative examinations of the micrographs that provide information regarding the results of an experiment in the form of visual impressions.

Direct Measurements. Direct measurements that can be made with the electron microscope are limited to those of size and shape, in other words, to the simple measurements of lengths and angles. Very often size is an important criterion in an experiment and in that case the electron microscope is invaluable. The actual measurement, of course, is made with some type of linear scale on the electron micrograph, the magnification of that micrograph being used to convert the measurements into the actual dimensions of the original object. It is, of course, necessary to know the magnification. Usually the electron microscope has been calibrated at a previous time by the research worker or by the electron microscopist so that the calibration is available. Furthermore, even if the accuracy of the calibration is unknown, the magnification can be depended upon to remain constant over a relatively long period. Hence, relative measurements can be made with high accuracy. Absolute measurements, when necessary, can be made to within 2%. The methods for obtaining accurate calibrations are described in the literature (2).

When it is necessary to make measurements to an accuracy that approaches the resolving power demonstrated in the electron micrograph, it is desirable that the worker have some understanding of those image phenomena that may influence his measurements. While it is not possible to describe these in detail here, it is desirable to call attention to the fact that the image contour phenomena that exist

in the electron microscope image vary greatly with focus, and in so
doing change the apparent size of small particles appreciably (*6*).

Other Information in the Image. At present most research
with the electron microscope depends on its use as an aid to our
vision. In other words, an experiment is performed and the result
carefully but qualitatively observed with the help of the electron
microscope. An obvious example of such an experiment is the fol-
lowing of the growth of bacteria. The method used may be a time
sampling of a large number of preparations in which the organisms
are made to grow directly on the supporting membrane (*7*). In the
electron microscope the obvious development of the colony can then
be followed, also the time interval of the observations can be narrowed
down so that the morphological changes that occur during the life
cycle of the individual cells can also be observed. Once the investi-
gator has become familiar with these changes he can then extend his
experiment further; for instance, by treating the cells of a certain age
with different chemicals in an effort to observe qualitatively the effect
on the structure in an indirect effort to determine the chemical changes
actually going on in the life process.

It is obvious from this rather brief example that the information
obtained will be intelligible if the experiments are properly designed
but will be in a somewhat intangible form that defies simple char-
acterization. The realization that the electron microscope is merely
a device to permit the brain to obtain visual information from the
microscopic world is probably the biggest step in the education of an
electron microscopist. Once this realization has come to the in-
vestigator, he will no longer have difficulty dismissing the complex-
ities of the electron microscope from his own problem.

D. LIMITATIONS OF ELECTRON MICROSCOPY

1. Instrumental Limitations

The electron microscope has two fundamental limitations:
resolving power and image contrast. The first of these, *i.e.*, the
limiting resolving power of the electron microscope, does not need to
be considered in detail in this chapter since it has only a slight bearing
on the results obtainable with the electron microscope in biological
problems. The latter, on the other hand, is quite important since
it is one of the most important instrumental factors determining

the amount of information the electron microscope can provide in any given problem.

Under ideal conditions, including perfect operation of the instrument and with the most suitable specimens, present day electron microscopes can be expected to yield a resolving power approaching 20 A. (2 mμ). In a few special cases instruments have been built that, under these ideal conditions, will provide resolving powers approaching the theoretical limit of the lenses in use at the present time—approximately 10 A. Nevertheless, the resolution actually demonstrated in an electron micrograph of a biological specimen may be considerably worse than this; it could be as much as 5 mμ for viruses and even as much as 50 mμ for internal structure of some of the larger bacteria. This is due to the fact that the observation of the structure of the specimen in the electron micrograph is dependent upon that structure producing differences in intensity in the image. Such differences in intensity arise only when there are differences in the electronic scattering power of adjacent portions of the structure in the specimen. Unfortunately, all organic materials are composed of light elements—mostly carbon, oxygen, and nitrogen—have approximately the same density and, hence, show very uniform scattering power for the electron beam. In a biological specimen it is possible to have a complex structure of different chemicals that nevertheless appear homogeneous in the electron microscope image. Thus, very often the electron microscope investigations of biological materials involve the development of techniques for introducing contrast in the image. For instance, as might be expected, heavy metals or compounds containing heavy metals will scatter electrons considerably more than organic materials and, hence, there exists the possibility of using such chemicals as a means of selectively staining the structures to be examined. Considerable preliminary work has been done on the use of such stains (*e.g.*, phosphotungstic acid or osmic acid vapor), though the method has by no means been fully exploited as yet.

There is also an instrumental method for increasing contrast, which is analogous to using a diaphragm in the objective of the light microscope. It consists of inserting a small opening in the objective lens of the electron microscope so that the effect of the differences in the scattering powers of various portions of the specimen are enhanced. At the present time the theoretical expectations of this method have been only partially achieved in practice, mainly due

to the considerable reduction in resolving power that accompanies
the introduction of an aperture.   In the electron microscope the
wavelength is so small that the numerical aperture can be as small
as 0.003 before there is any impairment of resolving power due to
diffraction effects.   The observed loss in resolving power is the result
of a technical difficulty of manufacturing a sufficiently small aperture
and having its surface uniformly conducting electrically.   There is,
in this regard, the additional problem of contamination of the aper-
ture due to electron bombardment and the subsequent charging of
the resulting semi-insulating coating.

2. Limitations Due to Preparative Techniques

In the use of the electron microscope in the field of biology, by
far the most important limitation of the method is the set of condi-
tions imposed on the preparation of the specimen.   Of these, the most
important is the fact that the specimen must be dried.   Next in
importance is the fact that the specimen must be prepared so that it
is less than a quarter of a micron in thickness.   Finally, the specimen
must be able to withstand high vacuum and often high temperatures
due to the electron bombardment.

Desiccation Artifacts.   Any microscopist interested in biological
materials is familiar with the disastrous distortion that many speci-
mens undergo in the final stages of drying.   At the same time he is
aware that under some conditions it is possible to dry biological ma-
terials without any apparent change as far as the light microscope is
concerned.   Generally speaking, the electron microscopist has not
been quite as aware of the serious nature of this problem because the
objects that tended to show the maximum distortion on drying were
in general objects that were too large for the electron microscope in
any case.   At the same time, studies on the size and shape of viruses
under natural conditions have agreed remarkably well with those
obtained with the electron microscope in the dried condition, so that
the electron microscopist is reasonably certain that for objects of the
size of viruses drying appears to have little, if any, effect.   Inter-
mediate between these extremes lie the bacteria, which to the casual
observer appear reasonably normal in the electron micrographs.
Critical bacteriologists, on the other hand, have realized that most
of the published electron micrographs of bacteria have shown evidence
of considerable shrinkage of the protoplasm away from the cell wall.

While this problem of distortion due to drying still remains the major limitation of the electron microscopy of biological materials, there are indications in recent work that the difficulty is not insurmountable. For instance, in tissue culture preparations, Porter and others (9) have greatly decreased the difficulties by careful fixation with osmic acid vapor. In the case of bacteria, recent work has indicated that the main reason that a large amount of shrinkage occurs on drying is the fact that the organisms were previously removed from their natural surroundings and suspended in distilled water. This, as might be expected, appears to wash away one or more layers of material that normally surround the cells, changes the concentration of salt in the cells through dialysis, and probably causes some other change due to osmotic effects. When organisms are grown directly on the supporting membrane and fed only by diffusion through that membrane it has been found that there is very little indication of artifacts due to drying. In this case the only contact that the preparation has with distilled water is during a brief interval in the transfer of the collodion bearing the cells from the agar plate to the wire screen (7). Since a plastic membrane separates the organisms from the water, there is no disturbance of the organisms. The effects of surface tension and the tendency of the cells and associated material to go into suspension are completely eliminated. Any changes caused by the contact with water must occur by diffusion through the membrane—a relatively mild effect.

It appears at the present time that there is no obvious fundamental reason why the geometrical structure or the morphology of organic materials cannot be preserved through the drying process. It is also quite apparent that considerable research must be carried out to understand the action of drying and to determine the best technique to be used. Certainly the progress made since the careful study of this problem was initiated (within the last five years) has been very encouraging.

Effect of Vacuum and Bombardment. The major distortion of the specimen occurs on the initial drying in air. It has been well established (8) that except in special cases there is no further change due to placing the dried specimen in a high vacuum. It has also been established that the bombardment by the electron beam also produces no further change, except in extreme cases. It is known, for instance, that if the specimen is subjected initially to an extremely rapid increase in electronic bombardment it is possible for large magnitude

changes to occur in the specimen, presumably due to a release of
bound water that occurs at high temperature and hence, simply
explodes the specimen by steam pressure. A slow increase of elec-
tronic bombardment on the same specimen, however, will seldom
show any effect until extreme intensities are reached. Even in this
case, the change usually is one of slight shrinkage accompanied by a
slight increase of the transparency of the specimen to the electron
beam. It appears that in this case the bound water is released suf-
ficiently slowly so that there is no rupturing of the specimen.

Generally speaking, extremely high intensity bombardment of
biological specimens is to be avoided for a number of other reasons
including the rupturing of the supporting membrane and the sub-
limation of salts and possibly other materials in the specimen.

Thickness of the Specimen. One of the factors restricting the
range of specimens that can be examined with the electron micro-
scope is the thinness required. In the case of examination of tissue
the problem of obtaining cut sections 0.2μ in thickness is a difficult
one. First there is the problem of actually cutting such a section.
A reasonably satisfactory solution appears to have been supplied by
Pease and Baker (*18*), as already mentioned. The problem of ade-
quate fixation still remains. However, with the means for cutting
the sections and the means for observing them at hand, workers in-
terested in this problem will be well equipped to discover a solution.

E. OUTSTANDING ACCOMPLISHMENTS OF THE
ELECTRON MICROSCOPE

In introducing the electron microscope to their fellow scientists,
it was natural that the early workers, including the writer, should
speak and write enthusiastically of its research possibilities, even
though many of the statements were based more on intuition and faith
than on available scientific evidence. That less than ten years later
the selection of the most important contribution of that instrument
has become a difficult task is probably the most outstanding aspect
of its accomplishments. Future possibilities estimated on the basis
of present work make the predictions of the early workers appear
conservative in the extreme.

In discussing the accomplishments of the electron microscope in
biology, a number of different researches could be described in con-
siderable detail. In the present work, however, it seems desirable

to describe only the application of the electron microscope to the
virus problem, since it is in this problem that the most widespread
intensive studies have been made and a definite measure of success
has been obtained. Nevertheless this section would be very incom-
plete if some of the other research projects undertaken were not men-
tioned.

The study of animal and insect tissues has been initiated and in
the case of muscle and other protein fibers the work of Schmitt and
co-workers (10) has been particularly noteworthy. Porter (9) on
the other hand, has started the study of the structure of animal cells
by means of a tissue culture technique. Nothing has as yet been ac-
complished in the study of histological sections because the necessary
techniques are still in the early stages of development.

In bacteriology the electron microscope has been used extensively.
However, a large part of the early work was purely exploratory in
nature and as a result there was very little, if any, organization of
the results. Mudd and Anderson (11) with a number of collaborators
have made an initial survey of the morphological characteristics of
different types of bacteria, which is the most complete at the present
time. G. Knaysi and co-workers, and more recently C. Robinow,
have undertaken a more careful study of the cytology of bacteria.
New preparation techniques developed recently by the author and
his co-workers are expected to greatly facilitate the latter work by
the elimination of many of the artifacts that troubled the early work-
ers. At the present time the future use of the electron microscope
in bacteriology shows great promise.

The Virus Problem

As soon as the electron microscope had reached the point where
it was established that high magnification images could be obtained,
it was applied to the examination of various types of viruses. The
main part of the early work was concerned with the verification of
the dimensions of the particles as obtained by statistical measure-
ments, i.e., by electrophoresis, diffusion, and ultracentrifugation. At
this point, its most important contribution was the accurate demon-
stration of the shapes of the viruses. While the indirect methods,
including in particular flow birefringence measurements, have been
able to differentiate qualitatively between rodlike and spherical par-
ticles and provide some estimate as to ratio of width to length, it

remained for the electron microscope to provide accurate pictures
of the shapes of the particles. Since most of the smaller viruses are
spherical, the electron microscope was not able to contribute very
much beyond this verification of shape and size. In fact, the statis-
tical methods of working with these viruses were more accurate and
at the same time permitted the study of the viruses under more nor-
mal conditions.

In later work, however, this situation changed considerably.
For instance, Wyckoff (12,13) and co-workers have studied the de-
tails of the crystallization of spherical virus particles. By evapo-
rating aqueous solutions of purified plant viruses, Wyckoff found that
the elementary particles were left in quite regular arrays over small
areas. The arrangement was the normal close packing that one
would expect from spherical particles. However, when he studied
replicas of the surfaces of small crystals of the elementary particles,
the crystals being formed from concentrated salt solutions, he found
that the arrangement was quite different. In fact, the same particles
now became arranged in a rectangular network and, moreover, the
individual particles were seen to be considerably distorted. Such
results could hardly have been obtained by any other method than
by the electron microscope and they have obvious extensions in the
study of the structure of other protein or macromolecular substances.

Another problem that recently has been given considerable
attention is in connection with the length of the tobacco mosaic
virus particles. It has been known for some time that old purified
solutions of tobacco mosaic virus have particles that show a rather
wide range of lengths. On the other hand, the particles examined
in dried films of the juice of a hair cell from an infected leaf are
remarkably uniformly 280 mμ in length, a result which Oster and
Stanley (14) consider conclusive evidence that the infective unit of
the tobacco mosaic disease is a particle 15 mμ wide and 280 mμ
long. To obtain the distribution of lengths for a large number of
particles in a given preparation is a direct and relatively simple matter
with the electron microscope. To obtain similar measurements in-
directly and by statistical methods is, of course, exceedingly difficult
if not impossible.

In the virus problem some of the most interesting work has been
done in connection with the infection of bacteria by the bacteriophage.
Here the electron microscopist has been fortunate because many of
the known bacteriophages have a very characteristic morphology

consisting of a relatively dense head between 50 and 100 mμ in diameter, though it is not always spherical, and a tail, which is usually 60 to 100 mμ long. The fact that most bacteriophages had such a structure could never have been proved definitely with the existing indirect methods of investigation. With the electron microscope a single direct examination of the particles in a purified suspension was all that was necessary. On the other hand, the fact that the bacteriophage particles have such a characteristic morphology has made them ideal subjects for electron microscope examination since they can be identified wherever they occur. There seems to be no other type of particle or material in their size range that has a similar structure.

The first examination of bacteriophage was made by H. Ruska in 1940. However, he had rather poor preparations in which there was a large amount of salt. The first clear observation of the bacteriophage was made by Luria and Anderson in 1942 (15). The latter workers were also able to show the adsorption of the bacteriophage particles on the whole cell and to show the ultimate lysis of the cell with the production of new bacteriophage particles (16)

Since that time progress has been rather rapid. Bacteriophage particles for staphylococci and actinomyces as well as a number of different strains for *Escherichia coli* have been examined with the electron microscope. A concentrated effort is being made to observe the mode of growth of the bacteriophage particles in the bacterial cell. The operation of the instrument is now better understood, with the result that it is possible to adapt it to obtain much better images of internal structures of the bacteria. In addition to this, the techniques for preparing bacterial specimens have been improved, so that it is now possible to examine the organisms at any definite time after infection by the bacteriophage, and under such conditions that there has been no disturbance of the specimen and that no material of interest in the specimen is lost or displaced. (See Figures 6–8, made in collaboration with Dr. Stuart Mudd, University of Pennsylvania.)

At the time of writing, there appear to be no fundamental technical limitations making it impossible to observe the complete life history of the bacteriophage particle in the bacterial cell, and, unless some difficulty not anticipated should appear, there seems to be no reason for not expecting this important problem to be solved in the near future.

References

GENERAL REFERENCES

1. Prebus, A. F., in *Colloid Chemistry*, Vol. V, J. Alexander, ed.   Reinhold, New York, 1944.
2. Zworykin, V. K., G. A. Morton, E. G. Ramberg, J. Hillier, and A. W. Vance, *Electron Optics and the Electron Microscope*.   Wiley, New York, 1945.
3. Burton, E. F., and W. H. Kohl, *The Electron Microscope*.   2nd ed., Reinhold, New York, 1946.
4. Zworykin, V. K., and J. Hillier, in *Colloid Chemistry*, Vol. VI, J. Alexander, ed.   Reinhold, New York, 1946.
5. Cosslett, V. E., *Introduction to Electron Optics*.   Clarendon Press, Oxford, 1946.
5a. Wyckoff, R. W. G., *Electron Micrography: Technique and Applications*. Interscience, New York, 1949.

SPECIFIC REFERENCES

6. Hillier, J., and E. G. Ramberg, *J. Applied Phys.*, **18,** 48 (1947).
7. Hillier, J., G. Knaysi, and R. F. Baker, *J. Bact.*, **56,** 569 (1948).
8. Hillier, J., and R. F. Baker, *J. Bact.*, **52,** 411 (1946).
9. Porter, K. R., A. Claude, and E. F. Fullam, *J. Exptl. Med.*, **81,** 233 (1945).
10. Schmitt, F. O., *Harvey Lectures*, Series XL (1944–45).
11. Mudd, S., and T. F. Anderson, *J. Am. Med. Assoc.*, **126,** 561, 632 (1944).
12. Price, W. C., and R. W. G. Wyckoff, *Nature*, **157,** 764 (1946).
13. Markham, R., K. M. Smith, and R. W. G. Wyckoff, *Nature*, **159,** 574 (1947).
14. Oster, G., and W. M. Stanley, *Brit. J. Exptl. Path.*, **27,** 261 (1946).
15. Luria, S. E., and T. F. Anderson, *Proc. Natl. Acad. Sci.*, **28,** 127 (1942).
16. Luria, S. E., M. Delbrück, and T. F. Anderson, *J. Bact.*, **46,** 57 (1943).
17. Williams, R. C., and R. W. G. Wyckoff, *J. Applied Phys.*, **17,** 155 (1946).
18. Pease, D. C., and R. F. Baker, *Proc. Soc. Exptl. Biol. Med.*, **67,** 470 (1948).

ACTION SPECTRA AND ABSORPTION SPECTRA

HAROLD F. BLUM, *National Cancer Institute and Princeton University*

A. INTRODUCTION

Photochemical reactions taking place in living systems underlie a number of phenomena, which may be classed together as *photobiological processes*; some are of the greatest importance to the living organism, others relatively trivial. Action spectra and absorption spectra, to the consideration of which this chapter is devoted, are useful tools in the study of the photochemical reactions concerned in these processes. To understand their use only the elements of photochemistry need be understood, although the photochemical reactions themselves may be quite complicated. The first step in any photochemical process, *in vitro* or *in vivo*, known as the *primary reaction*, is represented by:

$$A + h\nu = A^* \tag{1}$$

where A is a molecule of a given substance that will be referred to as the *light absorber*, A^* is the same molecule after it has been raised to an excited state by the capture of a quantum of radiant energy $h\nu$, h is Planck's constant (6.6236×10^{-27} erg sec.), and ν is the frequency of the radiation. The frequency is related to the wavelength λ as follows:

$$\nu = c/\lambda \tag{2}$$

where c is the velocity of light (3×10^{10} cm. sec.$^{-1}$). The following calculation, made for the wavelength 500 mμ ($= 5.0 \times 10^{-5}$ cm.), shows that the quantum has the dimensions of energy and indicates its magnitude:

$$h\nu = h(c/\lambda) = 6.6236 \times 10^{-27} \text{ erg sec.} \frac{3 \times 10^{10} \text{ cm. sec.}^{-1}}{5.0 \times 10^{-5} \text{ cm.}}$$

$$= 3.97 \times 10^{-12} \text{ erg} = 2.48 \text{ electron volts}$$

It is noted that the quanta dealt with in photochemical processes are much smaller than those of the ionizing radiation, which has been treated in other chapters. Most photobiological processes involve the action of many quanta, but this is not universally true, as illustrated in the values collected in Table I. When, as in scotopic vision of the human eye, only a few quanta are required to elicit a response, one must be very careful indeed about stray light; whereas in other cases one can be less exacting. Also, it is noted that in destructive processes, such as damage to cells by ultraviolet radiation, the cell may absorb many quanta of radiation without being severely damaged.

The primary reaction (1) is the initial event in any photochemical process, that is, a molecule must absorb a light quantum in order to reach the excited state from which it may proceed to take part in a chemical reaction. The reaction may vary from direct splitting of the light absorber molecule to long chain reactions (4,9). A treatment of the reactions that may follow the primary event is beyond the scope of this chapter, and for the moment it is only important to note that photochemical reactions of any kind depend upon the absorption of radiant energy, a fact often referred to as the first law of photochemistry, or sometimes, the Grotthus-Draper law.

Molecules in solution absorb only quanta of certain size ranges, determined by their internal structure.   Hence a given light-absorbing substance can only participate in a photochemical reaction when exposed to quanta of the appropriate size capable of bringing about excitation of its molecules.   This means that only particular wavelength regions can cause a given photochemical reaction, quantum size being inversely proportional to wavelength (equations 1 and 2). Thus, a means of characterizing any photobiological process is provided, because a given process is brought about only by those wavelengths absorbed in the primary reaction (1) by the specific light absorber.   This is the basis for the use of absorption and action spectra to identify the light absorber in a photobiological process, which is obviously one of the basic pieces of information required in the elucidation of any photobiological mechanism.   While the elementary principle is simple enough, the actual determination of the absorbing and exciting wavelengths is less so, since there are various complicating factors that may lead to incorrect interpretation of the measurements obtained.

TABLE I

Number of Quanta Required in Some Photobiological Processes

Photobiological process and source of data	Quanta required[a]
Photosynthesis: formation of 1 g. carbohydrate by the green plant. Based on quantum yield of approximately 1/12; see Franck and Gaffron, *Advances in Enzymol.*, **1**, 199–262 (1941)	10^{23}
Scotopic vision: minimal sensation of light in the human eye. Hecht, *Biol. Symposia*, **7**, 1–21 (1942)	5 to 14
Photodynamic action: hemolysis of the photosensitized rabbit erythrocyte. Blum and Gilbert (*2*, Ch. 7)	1×10^{10}
Destructive action of ultraviolet radiation: vesiculation of *Paramecium*. Giese and Leighton, *J. Gen. Physiol.*, **18**, 557–571 (1935)	1×10^{12}
Destructive action of ultraviolet radiation: inactivation of *Escherichia coli*. After Hollaender and Claus (*18*)	2×10^6
Inactivation of one molecule of an enzyme: trypsin. Verbrugge, *J. Biol. Chem.*, **149**, 405 (1943)	50^b
Inactivation of one molecule of a virus: tobacco mosaic virus. Uber, *Nature*, **147**, 148 (1941)	4×10^3

[a] All these values are based on estimates of number of quanta absorbed; most of these are approximate and they may vary to a certain extent with conditions.
[b] Additional values for inactivation of enzymes are summarized by A. D. McLaren, in *Advances in Enzymology*, Vol. IX, Interscience, New York, 1949, pp. 75–170.

B. ABSORPTION SPECTRA

An absorption spectrum represents the absorption of light by a given substance with respect to wavelength or frequency.   Examples are shown in Figures 1, 2, 5, and 9.   The absorption spectrum reflects the chemical composition of a substance, being determined both by the kind of atoms of which the substance is composed and the spatial and energy relationships of these atoms.   Thus, each chemical compound has a characteristic absorption spectrum, which may be used oftentimes to identify the compound or its general type, since similar
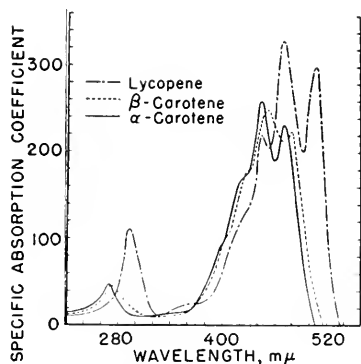
Fig. 1.   Absorption spectra of carotinoid pigments (20a). The "specific absorption coefficient" is ϵ/molecular weight.
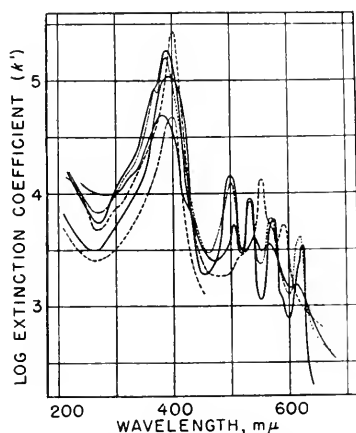
Fig. 2.   Absorption spectra of some porphyrins (2).   Plot of log k' permits ready comparison of shapes of curves.

molecules may have similar absorption.   The presence of certain radicals in large molecules, and their position in the molecule, may be reflected in the absorption spectrum, helping to determine chemical structure.

The results of absorption of quanta of radiant energy may be quite different, depending upon the size of the quantum and the nature of the absorbing material.   The effects of the large quanta constituting ionizing radiation, which may cause electrons to be removed from the atom or molecule thereby forming ion pairs, have been treated in other chapters and will not be discussed here.   This chapter is concerned with the absorption of small quanta, which results princi-

pally in raising the energy levels of electrons of the outer shells of the atoms.   The size of the quantum absorbed is equal to the difference in energy between the two levels concerned.

When absorption is by isolated atoms—*e.g.*, atomic hydrogen— there is a relatively small number of sizes of quanta that can be absorbed, because there is a limited number of energy levels that these electrons can occupy, and hence a limited number of possible transitions between these levels.   As a result the absorption spectra of atoms may consist of a relatively few isolated wavelengths.   Where the atoms are combined to form molecules the absorption spectrum is more complex, because the electronic energy levels of the constituent atoms are modified by the mutual relationships between the atoms, and, furthermore, the rotation of the atoms about each other and their vibration with respect to each other increases the number of quantized energy states.   The transitions between rotation states and between vibration states correspond to smaller quanta than those associated with changes from one electronic level to another.   Because of the large number and variety of differences in energy levels, the absorption spectra of molecules are composed of bands covering ranges of wavelengths instead of the monochromatic lines that characterize the absorption spectra of atoms; and in the case of highly complex molecules, such as are encountered in biological systems, the bands may form a continuum over a considerable range of wavelengths.   Examples of such continuous absorption spectra are shown in Figures 1, 2, and 9.   With proper conditions, *e.g.*, at low temperatures, it is sometimes possible to resolve such spectra into discrete lines or bands corresponding to the individual quantized states that compose them; but, for general purposes, absorption spectra of biological substances may be regarded as being continuous over a characteristic range of wavelengths.

In simpler compounds, a considerable degree of success has been had in relating chemical structure and absorption spectra, certain structures corresponding to absorption in particular wavelength regions, as might be expected from the relationship of quantum states and absorption.   However, our knowledge in this respect is still so incomplete that exact correlations between structure and absorption cannot always be made, and this is particularly true in dealing with biological substances.

It is customary to break up into three general divisions the spectral range considered in this chapter: (*1*) the visible, that region detected

by the human eye—wavelength approximately 390 to 650 mμ; (2) the
ultraviolet, that region of wavelengths shorter than the visible and ex-
tending down to the longer X-rays; (3) the infrared, that region longer
than the visible and extending as far as the shortest radio wavelengths.
The division into visible, ultraviolet, and infrared is arbitrary, since
the sensitivity of the human eye does not end abruptly at the points
assigned (see Figs. 3 and 4), and these divisions of the spectrum have
no true physical or biological meaning. Since room air absorbs virtu-
ally all ultraviolet radiation shorter than about 180 mμ and infrared
radiation longer than about 30 μ to such an extent that special meth-
ods of working are required, and also because the absorption by living
organisms themselves sets an even narrower range (see 3) most meas-
urements on biological systems are limited to these extremes. There
are other reasons for limiting the range of biological studies. Wave-
lengths lying in the infrared beyond about 1 μ represent small quanta,
which correspond only to the quantized states of rotation and vibra-
tion of the atoms in the molecule, changes in electronic energy level
requiring larger quanta. Longer wavelengths than this produce no
photochemical or photobiological action, because excitation of the
molecule that will permit it to take part in chemical reaction requires
such a transition in electronic energy level. Thus, so far as biological
action spectra are concerned, there is a wavelength limit, somewhere
in the near infrared, beyond which no such effects are to be expected.
The more distant region of the infrared in which only rotation and
vibration spectra are represented is of interest in another way, because
such spectra represent intramolecular forces characteristic of the
structure of the molecule. Their exploration with modern instru-
ments is becoming more and more useful in the study of the structure
of organic compounds, but this is a subject outside the scope of this
chapter (see 5,8).

1. Quantitative Measurement of Absorption Spectra

The complete determination of an absorption spectrum requires
measurement of the relative extent of absorption at various wave-
lengths. Methods of obtaining and of measuring monochromatic
radiation are treated very briefly in Section C1; here only the meas-
urement of absorption of monochromatic radiation is discussed.

Absorption Laws and Units. It is essential to consider first the
so-called absorption laws. The intensity of a parallel beam of

monochromatic radiation may be defined as the number of quanta incident in unit time upon unit area of a surface normal to the beam (see also last paragraph of this subsection). If the beam enters a homogeneous absorbing medium, a given fraction of the incident quanta is absorbed in passing through a given thickness of the medium, the fraction being determined by the number of absorbing molecules presented by the medium per unit thickness. This means that intensity is always decreased by a given amount in passing through a given thickness. Thus, if dl is an infinitely small unit of thickness, and dI is the change in intensity of a monochromatic beam in passing through this thickness:

$$dI/dl = -kI \tag{3}$$

where I is the intensity at any point and k is a constant. Integration of equation (3) yields:

$$\ln (I/I_0) = -kl \tag{4}$$

or, in exponential form:

$$I = I_0\, e^{-kl} \tag{5}$$

where I_0 is the intensity of the incident beam, and I is the intensity after having passed through thickness l. The constant, k, in equations (3–5) is usually called the *absorption coefficient*.

Either equation (4) or (5) is an expression of the Bougeur-Lambert absorption law, which is often written with common logarithms and in the reciprocal form

$$\log (I_0/I) = k'l \tag{6}$$

where k' is a constant, usually called the *extinction coefficient*. When based on the same unit of path length, l, $k' = k/2.303$. The corresponding exponential form is sometimes used

$$I_0/I = 10^{k'l} \tag{7}$$

The relationship may be found expressed in any of these forms (equations 4–7).

Since the fraction of the quanta absorbed is directly proportional to the number of absorbing molecules, we may also write:

$$\log (I_0/I) = \epsilon C l \tag{8}$$

where C is the concentration, and ϵ the *molecular extinction coefficient*. This is the common form of expression of Beer's law, which is the general basis for absorption spectrophotometry of solutions. Another term in common use is the *density*, D, which is defined as:

$$\log (I_0/I) = D \tag{9}$$

The length of the absorbing path is not specified. In equations (3-8) the numerical value of k, k', or ϵ depends upon the unit in which the path length, l, is measured; this is usually the centimeter. The *molecular extinction coefficient* is generally defined as the value of D a solution must have if the concentration is 1 mole per liter, when the absorbing path is 1 centimeter.

The nomenclature for expression of the absorption laws is by no means uniform, nor are the symbols or the units. It is therefore necessary in absorption spectroscopy to understand the basic relationships clearly, and to ascertain the units used by a particular author; it is equally important to be sure that your own results are clearly and completely expressed. While the Bougeur-Lambert law holds for any homogeneously absorbing medium, Beer's law sometimes fails when applied over wide ranges of concentration. Thus, it is well to state the approximate concentrations at which measurements are made.

The numerical values of k, k', or ϵ apply to monochromatic radiation of a specific wavelength for which they have been determined. Hence an absorption spectrum may be expressed by plotting k, k', or ϵ against the wavelength measured in $m\mu$ (Figs. 1 and 2) or some other unit. Sometimes, however, they are plotted against the frequency ν in sec.$^{-1}$ or against the wave number ($= 1/\lambda$, when λ is measured in centimeters). For purposes of comparison it is sometimes convenient to plot the logarithm of the absorption coefficient as in Figure 2; this is particularly useful when dealing with solutions of unknown concentration.

In biological experiments, the concentration of a substance is often unknown, and hence data cannot be expressed in terms of molecular extinction coefficients or Beer's law; but absorption coefficients, extinction coefficients, or, in some cases, density may be used. In much of the discussion of action spectra that follows, it makes no difference which of these units is used, because only relative values are important.

The intensity, I, was defined above as the number of quanta per unit area per unit time. Measurements of radiant energy are usually

made with a thermocouple or bolometer, or are referred to these in-
struments, which measure radiant energy flux per unit area in terms of
energy units.   Hence it is general practice to express intensity in
terms of energy units per unit area per unit time;   and herein I'
will be used to designate the intensity when expressed in these units.
Since the absorption coefficient is determined from the ratio of intensi-
ties, the numerical value is the same whether I or I' is used, i.e.,
numerically $I_0/I = I_0'/I'$.   However, when one wishes to compare
the effectiveness of radiation of different wavelengths, as in dealing
with action spectra, the definition of these terms becomes important
(see Sect. C2).

Measurements.   In biochemistry, absorption spectroscopy is
applied extensively in the identification and estimation of compo-
nents of biological systems; in most cases the material is removed
from the living system, and its absorption spectrum measured in
some solvent.   The absorption spectrum of a substance may differ
considerably according to the solvent in which it is dissolved, and
identification is therefore made by comparison of the absorption spec-
trum to those of known compounds in solution in the same solvent.
Since certain structures have characteristic absorption, such studies
may assist in determining the chemical structure of substances of un-
known constitution although advances in this direction are somewhat
limited (5,7,24).   The determination of absorption spectra under
conditions where one deals with homogeneous substances in known
solvents and where scattering of light is usually not an important
factor is quite straightforward, since appropriate equipment is avail-
able for this type of measurement (see 5,8,9a).   Thus in biochemistry,
methods of extraction and problems of organic structure may out-
weigh those of absorption spectroscopy.   The biophysicist and
physiologist, on the other hand, are more commonly presented with
the problem of measuring absorption spectra of substances inside liv-
ing systems, and here new factors enter, for example, light scattering.
Particular difficulties are encountered in attempting to determine the
light absorber in a photobiological process, where the action spectrum
is compared to the absorption spectrum of a substance that must,
as a rule, be removed from the living system to an entirely different
environment, i.e., to some solvent that does not resemble closely the
milieu of the living organism.   Before considering some of these dif-
ficulties, methods used for measurement of absorption spectra will be
discussed briefly.

Most measurements of absorption spectra are made in solutions that do not scatter light to an appreciable extent, and most modern methods are adapted to this particular type of measurement. Such methods have been adequately treated in various places (*e.g.*,*5*,*8*,*9a*), and hence only a brief discussion of the general principles of absorption spectroscopy is needed here. The measurements consist essentially in determining for each wavelength the ratio I_0/I for a sample of the material of known thickness, l; from this information the absorption coefficient or extinction coefficient may be calculated.

The most direct method—not commonly used now except for the rotation–vibration region of the infrared—is to measure the intensity of a monochromatic beam successively with and without the absorbing material in the beam; the former gives I, the latter I_0. The intensity may be measured with an instrument whose response is directly proportional to the intensity, for example, a thermocouple; or with an instrument in which correction can be made for any departure from such proportionality. Measurement in absolute units is unnecessary since only the ratio of intensities is required, and the instrument need not be nonselective as regards wavelength.

Another method employs the spectrograph. In this instrument the spectrum of some source, say an iron arc, dispersed by means of a prism or grating, is arranged to fall on a photographic plate. Photographs are made with and without the absorbing medium interposed, and the intensities of the various wavelengths determined by analysis of the photographic negative. This method has the advantages that it provides a permanent record, and that a wide range of wavelengths may be measured with a single pair of exposures. It has the disadvantage that the response of the photographic plate is not directly proportional to the intensity and hence correction must be made in terms of the blackening characteristics of the plate; this requires appropriate equipment and is time-consuming. In some of the older investigations this correction was not made, leading to serious errors in interpretation. Such mistakes are to be found in the biological literature somewhat later than in the physical.

The methods most in use at present employ spectrophotometry in one form or another. By splitting the monochromatic beam into two beams of equal intensity and interposing the absorbing material in one, the reduction in intensity in passing through this material may be determined by comparison of the two beams photometrically. This gives the ratio I_0/I directly without determining the

intensities separately, so that the measurement is made at a single reading. In the visual spectrophotometer the intensities of the two beams are adjusted until they match to the eye, whereas in photoelectric spectrophotometers, the difference in intensity of the two beams is measured by the differential response of photocells. By appropriate devices the photographic plate may also be used photometrically; two matched spectra are photographed—one with the absorbing material interposed, the other without—in such a way that the plate blackening in the two cases can be compared directly without correction for the characteristics of the plate.

Methods of absorption spectroscopy have been developed for use with the microscope, particularly the quartz microscope for the ultraviolet. Here the microscope is focused on a given object in the cell, and the absorption of a monochromatic beam measured by photoelectric photometry or in some other way. This method is limited to objects whose diameter is about four times the wavelength of the light employed or greater, and is a function of the angular aperture (6,25).

Materials in solution are usually contained in glass or quartz absorption cells having flat parallel walls placed normal to the beam in determining I; and the same or a matched container filled with the solvent is then used in determining I_0. In this way the ratio I_0/I is an index of the absorption of the dissolved substance, provided the solution is reasonably dilute; factors such as absorption by the solvent and reflection by the walls of the chamber cancel out, since they are the same in both determinations. In measuring absorption by a solid, say a glass color filter, I_0 is usually determined with no material interposed. In this way one obtains the absorption with respect to air, and may or may not take into consideration the reflection from the surfaces, depending upon the purpose of the measurement. In biophysics, various situations may arise in which it is desirable to treat the question of reflection in different ways.

2. Absorption Spectra of Substances in Living Systems

Living systems provide a complex environment in which it may be difficult to know the exact physical or chemical state of a given substance. The greater part of the living cell displays some degree of organization, i.e., the molecules are at any given moment distributed in some characteristic pattern, which varies in its degree of

exactness for different structures. In some parts of cells, and in some of the fluids outside the cell proper, there are substances dissolved in simple solution, but this is probably the exception rather than the rule. In such organized systems the quantum states of a given molecule may be different from those of a molecule of the same kind that finds itself in a simpler environment, say in an organic solvent. Close chemical combination with another molecule may alter some of the electronic energy levels, and even very loose association of molecules may be expected to alter their rotation and vibration states; changes in absorption spectra must result. Minor alterations, particularly of rotation–vibration states, would usually not be expected to affect the absorption spectra of large molecules to a great extent; and the general character of the absorption spectrum may often be retained. Thus, one may hope in many instances to find close resemblance between the absorption spectra of compounds extracted from living systems and the same compound inside the living cell, but one must also be prepared for discrepancies of considerable magnitude. Hydrogen ion concentration is particularly liable to affect absorption spectra of substances in solution, as witness the case of acid–base indicators, and hence, when the pH of the extracted solution differs appreciably from the *in vivo* condition, wide discrepancy in absorption spectra may appear.

Another factor that may lead to discrepancy is failure of Beer's law, *e.g.*, the absorption spectrum of some compounds in solution may show considerable shift in absorption spectrum with dilution. The causes for such changes are not clearly understood, and it is well to guard against disagreement by making absorption measurements of extracted substances in as nearly as feasible the same concentrations as those at which they exist in the living system concerned.

With the complex nature of the systems under consideration, it is impossible to make any general rule regarding the nature of the differences to be expected between the absorption spectrum of a compound inside a living cell and the same compound when extracted therefrom. This is a matter calling for the exercise of careful judgment in interpretation.

3. Effect of Light Scattering

All molecules scatter light to a certain extent; in general, the scattering increases with molecular size and with decrease in wave-

length of the incident light, but there are complicating factors, such as interference, which make this rule far from universal (see *1,10*). For particles with diameters near the wavelength of the incident light, the Rayleigh formula is obeyed, according to which scattering increases inversely with the fourth power of the wavelength. Scattering by larger particles such as are often encountered in living systems, *e.g.*, scattering by reflection and refraction at the boundaries of cells, is usually dependent to a much less extent upon wavelength.

Ordinary methods of absorption spectroscopy, like those described above, succeed because the systems dealt with do not scatter appreciably in the spectral regions studied. Scattering introduces error, when such methods are used, because the beam that enters the absorbing medium normal to its surface is scattered in all directions, more or less at random, so that much of the light does not reach the measuring element, which is arranged to receive an emergent beam normal to its surface. When scattering is high, only a small fraction of the light that has passed through the absorbing substance may reach the measuring element, and hence the estimated absorption is higher than the true absorption, since much of the transmitted light is not measured. The amount of loss by scattering must depend upon the geometric arrangement of the instrument, as well as upon the degree of scattering by the material. Thus, measurements of absorption by scattering materials obtained with ordinary instruments usually have little significance.

There are no standard methods for measuring absorption by scattering materials, and their development is to be desired. Several devices have been used. One is to diffuse the beam before it enters the absorbing material by interposing a diffusing plate, *e.g.*, a plate of quartz with a ground surface, so that the entering light is already completely scattered. This method uses ordinary equipment, but has the disadvantage that the intensity may be so much reduced that very sensitive measuring devices must be employed. Essentially the same result is accomplished in the microscope since the illumination enters the material on the stage from a wide angle so that this instrument is quite satisfactory for the determination of the absorption of various components of living cells. Another method is to place the absorbing material very close to a large flat-surfaced receiving element, or one made in the form of an integrating sphere. Thus the element receives most of the light scattered over $180°$ solid angle. Kirby-Smith (*20*) developed a method for studying the absorption of

separated epidermis in which the intensity was measured by an ele-
ment that could, by use of stops, be caused to accept the light from
various solid angles, extrapolation then being made to 180° solid
angle.   These methods have been used chiefly for measurement of
transmission (I/I_0) by a scattering layer, e.g., the human epidermis,
rather than for determining absorption by the layer.   For the latter
purpose one needs to take into account the light scattered throughout
a 360° solid angle, a more complicated measurement.

Attenuation Coefficient.   In a homogeneous scattering and ab-
sorbing medium, the integrated intensity decreases with the dis-
tance traveled, in the same manner as in a nonscattering medium.
Thus, one may write, in form comparable to the Bougeur-Lambert
law (equation 5):

$$I \,=\, I_0 \, e^{-al} \tag{10}$$

where a is a constant, known as the *attenuation coefficient*, which
recognizes both absorption and scattering.   In systems containing
no particles larger than molecules, a relationship corresponding to
Beer's law may be expected to hold, at least within limited ranges of
concentration.   Equation (10) should hold for systems in which
scattering is due to particles larger than molecular size, so long as
their size and distribution are homogeneous, but, unfortunately,
living tissues are seldom homogeneous throughout any great thick-
ness, and hence attenuation coefficients cannot be accurately cal-
culated.   The concept of the attenuation coefficient is, however, a
useful one in the idealization of such systems for theoretical considera-
tion (e.g., *14*).

4. Fluorescence

Many biological substances fluoresce, i.e., the molecule after ab-
sorbing a quantum emits within a very short interval of time an-
other quantum, usually smaller than that absorbed.   Hence a certain
portion of the absorbed radiation is re-emitted as radiation of longer
wavelength.   The direction of this fluoresced radiation, like scattered
radiation, bears no relationship to the direction of the incident beam.
Thus fluorescence introduces error of the same kind as scattering,
but in most cases the magnitude of loss due to this factor is not
great.

C. ACTION SPECTRA

The action spectrum for a photobiological process is an expression of the variation of the effectiveness of the exciting radiation with the wavelength; examples appear in Figures 3–6, 8, and 9.   There is no generally accepted way of expressing action spectra that corresponds to the use of the absorption coefficient in absorption spectra.   The reciprocal of the incident energy ($1/I'$) required to produce a given photobiological response is probably the expression most commonly employed (as in Figs. 4 and 5), but has numerous disadvantages and its use is by no means universal.   In fact, with regard to the whole
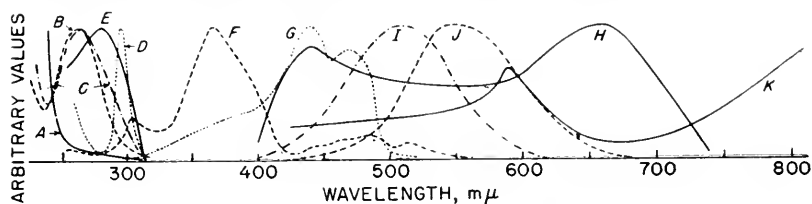
Fig. 3.   A number of action spectra (2).   The ordinates are I/I'_0 corrected to unity at the point of maximum action.   (A) Hemolysis of erythrocytes.   (B) Killing of a bacillus, *E. coli*.   (C) killing of a yeast, *Saccharomyces cerevisiae*.   (D) Erythema of human skin.   (E) Vesiculation of a protozoan, *Paramecium multimicronucleata*.   (F) Vision of an insect, *Drosophila*.   (G) Tropic bending of the oat coleoptile.   (H) Photosynthesis of the wheat plant.   (I) Scotopic vision (rods) of the human eye.   (J) Photopic vision.   (K) Photosynthesis by purple sulfur bacteria, *Spirillum rubrum*.   The curve reaches a maximum about 900 mμ.   The original references are cited in Blum (2).   Accuracy of some of these action spectra may be questioned; but wavelength limits are probably nearly correct, and this alone serves to illustrate spectral range of such effects.

subject of action spectra, the systems studied and the problems of measurement are so diverse that very few general rules can be laid down.   The purpose for which the action spectrum is obtained may dictate the mode of expression.   For example, in illuminating engineering, and in many other problems involving human vision, one is concerned with the amount of incident energy required to stimulate the visual response (Fig. 4), and this is usually measured in ergs or some other energy unit.   On the other hand, if one wishes to determine the nature of the light absorber, *i.e.*, the photosensitive retinal pigment in this case, a mode of expression is desirable that permits one to deal in terms of the number of quanta absorbed.   To estimate the absorption, indirect measurements are often required, sometimes

involving problems that are essentially biological rather than physical.

1. Sources and Measurement of Radiation

The exact determination of action spectra requires an appropriate source of monochromatic radiation. However, as may be judged from inspection of the action spectra shown in Figure 3, the purity (monochromaticity) of the monochromatic radiation required may depend upon the detail of the particular action spectrum and the
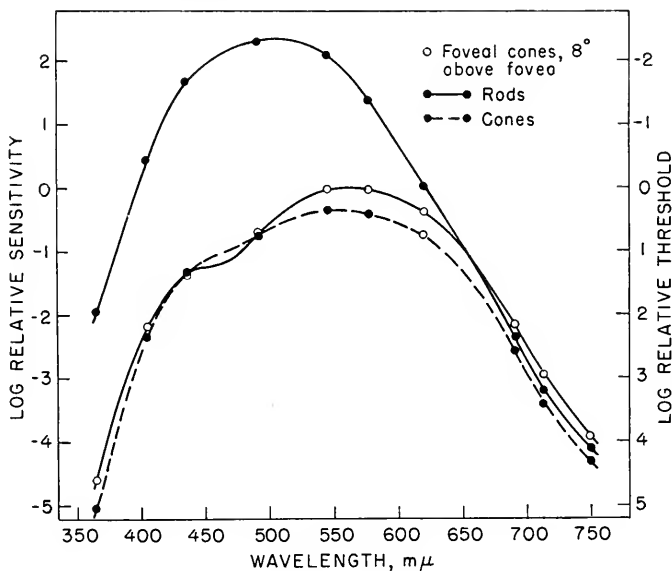
Fig. 4. Spectral sensitivities, measured as reciprocal of threshold ($1/I'_0$), for foveal and peripheral cones (photopic vision) and peripheral rods (scotopic vision) (28). All sensitivities are expressed relative to the maximum sensitivity of the fovea.

use to which the information is to be put. It may frequently be necessary to sacrifice purity in order to obtain enough intensity to elicit the photobiological process under study, for in some instances quite high intensities are needed. These are matters that cannot be determined without some preliminary information about the material to be studied. In some cases approximate measurements using polychromatic radiation may be as interpretable as those that can be obtained with monochromatic radiation.

Sources of radiation are of two types, presenting somewhat different problems with regard to the isolation of monochromatic radiation. "Line" sources, such as, for example, the mercury arc, emit only a limited number of wavelengths, which appear in the spectrum as discrete lines. The wavelengths correspond to the sizes of quanta emitted when electrons jump from higher to lower energy levels in the atom, and since the levels are characteristic of the emitting atom,
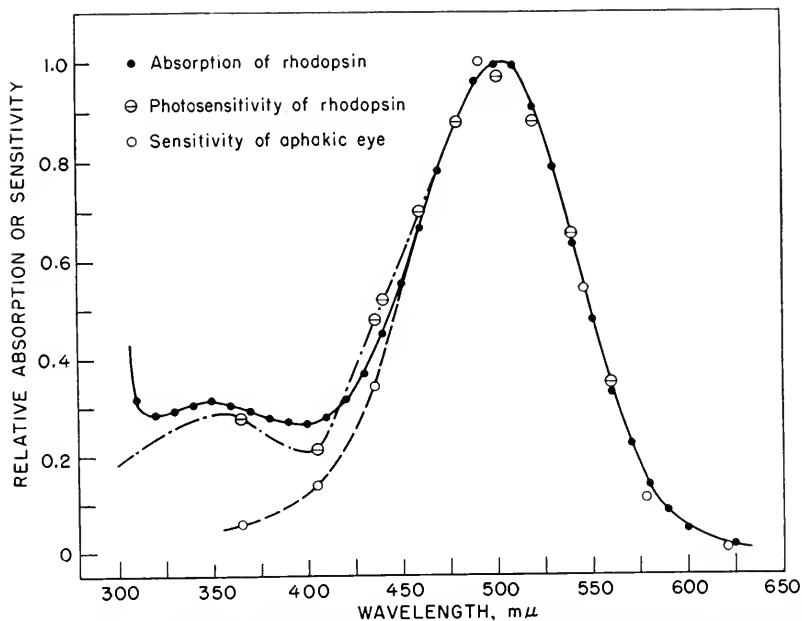
Fig. 5. Comparison of absorption spectrum of rhodopsin, spectral sensitivity of a solution of rhodopsin, and spectral sensitivity of scotopic vision in aphakic (lensless) subjects (*29*).

the wavelengths of the lines emitted by a given kind of arc are always the same, although their relative intensities may vary. With such sources it may be a relatively easy matter to isolate virtually monochromatic radiation by means of a monochromator or filters. The difficulty with such sources is that they provide only a limited number of wavelengths, often too widely separated to furnish adequate detail for an action spectrum. Actually, the only source of this type developed to a sufficient state of dependability for many biological studies is the mercury arc. This arc provides a sufficient

number of lines in the ultraviolet below 334 mμ for most purposes, but only scattered lines in the longer ultraviolet and visible. A mercury arc operated at very low pressure may emit most of its energy (about 85%) in the resonance line, 253.7 mμ; this is a useful source of nearly monochromatic radiation for some purposes, but is obviously of little use by itself for determining action spectra.

"Continuous" sources, of which the tungsten filament lamp is an example, emit all wavelengths over a wide spectral range. To isolate approximately monochromatic radiation from such a source requires a very efficient monochromator. Monochromators of prism type are the more commonly employed, but gratings that concentrate most of the energy in a single order may be satisfactory. The latter have the advantage that the dispersion is the same for all wavelengths. Increased purity may be obtained by combining two prisms or two gratings into a "double" monochromator. Even with the largest monochromators available the area that can be successfully illuminated is limited to a relatively small size, inadequate for many biological studies.

Filters, on the other hand, permit much larger areas to be irradiated, but the older types of "monochromatic" filters (e.g., glass or gelatin film) transmit wide spectral bands, and cannot be compared to monochromators as regards purity. However, they may, if properly chosen, serve well enough for isolation of monochromatic radiation from line sources when the lines are not too close together. During the past few years interference filters have been developed that isolate narrower spectral bands than is possible with the older types, and these will no doubt prove of considerable value for photobiological studies. As much as 45% of the incident radiation may be transmitted, the width of the transmitted band being usually less than 20 mμ, for half peak transmission. This makes them appropriate for use with line sources, although they may not provide sufficient purity for successful use with continuous sources. At present such filters are available only for wavelengths in the visible region, but they may soon be produced for the near infrared and ultraviolet. As is the case with older types, these filters may transmit elsewhere than in the principal band of transmission, and this possibility must be kept in mind. Further remarks about the use of filters will be found at the end of Section C3.

Generally, no matter what the object of the study, the principal physical measurement that has to be made is the determination of the

incident radiation. In practice it is usual to arrange the apparatus so that the measuring element can be interchanged with the surface of the biological material under study. The measuring element of choice is a thermocouple or a bolometer, since these are nonselective with respect to wavelength, but other, more sensitive instruments, e.g., photoelectric cells, may be used if their spectral sensitivity is known and can be taken into account. One usually measures the intensity of the incident beam, and then allows the beam to impinge upon the biological material for a certain length of time; the incident energy being determined in energy units per unit time (I'), the dose received by the biological system can thus be calculated in energy units. It is not strictly necessary to measure the intensity in absolute units, for in many cases only relative values are needed to interpret the results. However, important information can sometimes be gained by estimating the number of quanta absorbed, and for this purpose energy values are required in absolute units. Moreover, the reproduction of measurements by other investigators may be facilitated if the intensities are accurately known; so these values should be presented in absolute units when feasible.

2. Photochemical and Photobiological Aspects

Any photobiological process is, generally speaking, a photochemical reaction; it may be an involved one, in which some of the components are complex biological responses, based themselves on chemical reactions but about which relatively little is understood. In most cases the intermediate steps between the primary photochemical act, represented in equation (1), and the final measured response are virtually *terra incognita*, which one attempts to explore by various experimental means. The interest in action spectra focuses on the primary act—the absorption of quanta by the light absorber—and although, as a rule, this can be studied only in terms of the final response, one has the advantage that the intermediate steps can be disregarded as long as the relationship between these initial and terminal events is known.

Reciprocity Law. The simplest relationship is a direct proportionality between the number of quanta absorbed in the primary act and the final step in the photobiological reaction. In this case, a given number of quanta produce a given measurable end response,

and the process is independent of the intensity. This is generally expressed by the relationship:

$$\text{intensity} \times \text{time} = \text{a constant} \qquad (11)$$

which is called the *reciprocity law*, or sometimes the Bunsen-Roscoe law. It is obvious that such a relationship must hold for the primary act (equation 1) but it should be equally obvious that it may or may not hold for the over-all process, because, if any intermediate step is dependent upon the intensity, the end response must also depend upon the intensity, *e.g.*, a reversible intermediate reaction may depend upon the rate at which some substance is formed in the primary reaction, and hence upon the intensity. As a matter of fact, reciprocity is displayed by many photobiological processes over a wide range of intensities, and it is probable that numerous instances of failure that are described are apparent rather than real; but failures are nevertheless to be anticipated.

One cause of apparent failure of reciprocity is that the time required for a given biological end point to be reached includes the time required for processes set off by photochemical reactions, but which are not themselves dependent upon light. For example, photosensitized red blood cells hemolyze after sufficient exposure to light. If the exposure is terminated at the proper time, before the cells have begun to disintegrate, they will, nevertheless, hemolyze in the dark later on. Obviously, if one continues the exposure until, say, 50% hemolysis, one includes, in calculating the dose, a period during which the process is more or less independent of the light impinging upon the system. In this case departure from reciprocity is observed. On the other hand, if one exposes the cells to a small dose, such as will cause hemolysis to occur at some later time, say 24 hours after exposure, the magnitude of the dose can be accurately calculated, and when this is done reciprocity is found. The determinations in the latter case require that a series of like samples of cells be exposed to various doses, at various intensities, and then left for 24 hours, at the end of which time the minimum dosage that has produced hemolysis is determined (*2*, Ch. 4).

We have here an example in which the end point is a specific biological response set off by a dose of radiation delivered at some earlier time. This is a common type of criterion. Other examples are the killing of bacteria by ultraviolet radiation (*18,22*) and the induction of tumors by repeated doses of ultraviolet radiation, the

tumor appearing only after the lapse of several months during which many doses are delivered (12). In all these effects reciprocity holds at least over wide ranges of intensity, but the experiments must be properly designed to demonstrate this point.

In order to obtain an interpretable action spectrum, it is necessary in most cases to know that reciprocity is obeyed, or to know the intensity dependence so that appropriate correction can be made. The dependence upon intensity may also give important clues regarding other aspects of a photobiological mechanism, and hence the reciprocity relationship is one of the first that should be explored in the investigation of a photobiological process. It is possible in some instances, however, to compare minimal (threshold) stimulating intensities, as in investigations on the human eye, and so to obtain action spectra without reference to reciprocity.

Quantum Efficiency and Quantum Effectiveness. Measurements of intensity or dosage are usually made in terms of energy units, whereas photochemical, and hence photobiological, efficiency must be reckoned in terms of the number of quanta absorbed. The efficiency of a photochemical reaction is measured by the *quantum yield*, which is expressed as follows:

$$\text{quantum yield} = \frac{\text{number of molecules changed chemically}}{\text{number of quanta absorbed}}$$

The quantum yield for the primary act, as described by (equation 1) would be unity; but for a reaction involving several steps it may range from a small fraction to a value much larger than unity in the case of chain reactions. In photobiological processes the quantum yield can seldom be calculated, photosynthesis being an exception. Thus the term *effectiveness* is more appropriately applied to most photobiological processes than *efficiency*, the more specifically defined term.

In comparing the effectiveness of different wavelengths it is necessary to take into account the relationship between incident energy and incident number of quanta. From equations (1) and (2) we see that the number of quanta, N, is related to the energy, Q, as follows:

$$N = Q/h\nu = Q\lambda/(hc) \tag{12}$$

and since h and c are constants, we may write:

$$N \propto Q\lambda \tag{13}$$

We have defined I_0 as the incident intensity in terms of number of
quanta per unit time, and I_0' as the incident intensity in terms of
energy units per unit time. It follows from equation (13) that, for
unit time:

$$N = I_0 \propto I_0'\lambda \qquad (14)$$

In comparing the effectiveness of different wavelengths, when one
has no means of estimating the actual numbers of quanta—as is most
often the case—it is convenient to use either of the reciprocals,
$1/(Q\lambda)$ or $1/(I_0'\lambda)$ as the index of *relative effectiveness*. Over short
wavelength ranges this correction may be relatively unimportant
but, when the action spectrum covers a wide spectral range, as for
example in photosynthesis, considerable error may be introduced if
the correction is not made.

3. Estimating Absorption

In photochemical experimentation it is common practice to ar-
range the apparatus for "complete" absorption of the incident radia-
tion, that is, the dimensions of the reaction chamber and the concen-
tration of the light absorber are chosen so that the entering beam is
reduced to a negligible intensity in passing through the system.
Provided scattering is negligible or can be corrected for, the ab-
sorbed radiation can be, in such cases, taken as equal to the incident
radiation. This greatly facilitates the measurements; and efficiency
can be calculated in these terms. The success of this sort of arrange-
ment depends upon, among other things, the maintenance of vir-
tually uniform concentration of the reactants throughout the system,
usually by stirring. It is sometimes possible to use the same device
in studying a photobiological process, for example, in photosynthesis
by a microscopic organism such as the alga *Chlorella*, since the or-
ganisms, suspended in an aqueous medium, may be stirred in the re-
action chamber in the same way as more simple chemical reactants,
maintaining for them a homogeneous environment (*15*). Suspensions
of bacteria have been treated in the same way (*18*). In such cases
stirring must be adequate. The process of photosynthesis lends itself
further to study by photochemical methods, since one can measure
the oxygen production, from which may be calculated the amount of
carbon dioxide reduced to carbohydrate.

The situation is not quite the same when one measures photosyn-

thesis by the leaf of a higher plant, where stirring is not possible (compare Figs. 3H and 6). Here, one meets a situation comparable to that found in the majority of photobiological processes, where the photochemical reaction goes on in a given part of an organism, whose relationship with other parts is fixed, and one cannot assume that the incident energy is a measure of the absorbed energy, nor that the chemical environment is homogeneous. In systems of this type, it is possible when the conditions are appropriate to use an approximation
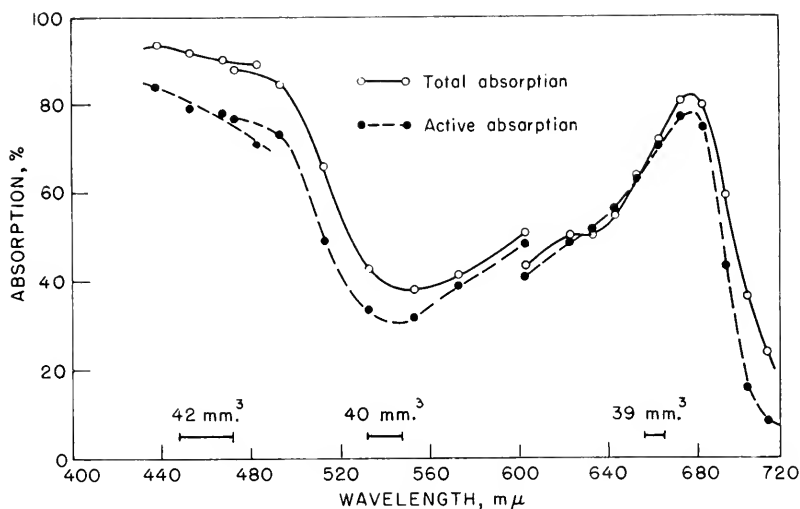
Fig. 6. A comparison of absorption by cells of *Chlorella* to "active absorption" (action spectrum) calculated on the basis of quantum yield 0.084 (17).

in estimating the effectiveness from the incident energy, or intensity required to induce a given biological end point. Let us assume an ideal photobiological process in which the reciprocity law is obeyed, so that the end point corresponds to the absorption of a given number of quanta. Let us write Beer's law in a slightly different form, by assuming that I and I_0 are measured over the same given interval, so that we deal with corresponding numbers of quanta instead of intensities:

$$N/N_0 = e^{-gcl} \tag{15}$$

where N_0 is the incident, N is the emergent number of quanta per unit area for a given interval of time, g is a constant corresponding

to the molecular extinction coefficient as ordinarily expressed in
equation (8), and C and l are, respectively, the concentration of the
light absorber and the thickness of the system.

Now, if N_a is the number of quanta absorbed by the light
absorber:

$$N_a = N_0 - N \tag{16}$$

We may substitute in (15) and rearrange:

$$N_a = N_0(1 - e^{-gCl}) \tag{17}$$

Expanding the exponential:

$$N_a = N_0 \left(gCl - \frac{g^2 C^2 l^2}{2!} + \frac{g^3 C^3 l^3}{3!} \ldots \right) \tag{18}$$

If the absorption is less than, say, 10%, no important error is intro-
duced by neglecting powers higher than unity, and:

$$N_a = N_0 gCl \tag{19}$$

Now, if C and l are constant, as may be assumed to be the case in
many photobiological processes, and if their values are such that ab-
sorption is small:

$$N_a \propto N_0 g \tag{20}$$

Since g is proportional to the absorption coefficient, k, the extinction
coefficient, k', and to the molecular extinction coefficient, ϵ, any of
these quantities can be substituted in the proportionality (20).

Assuming that the number of quanta required is the same for all
wavelengths, the relative effectiveness for a given wavelength may be
represented as the reciprocal of the relative number of quanta ab-
sorbed, so $1/(Q_0\lambda)$ or $1/(I_0'\lambda)$ may be plotted as the measurement of
effectiveness (equations 13 and 14), as has been done in Figures 5,
8, and 9. This permits a direct comparison with the absorption spec-
trum of a substance suspected of being the light absorber; the choice
of expression of the absorption being immaterial, since:

$$1/(Q_0 \lambda) \propto g \propto k \propto k' \propto \epsilon \propto D \tag{21}$$

We note that the analysis above is based on several assumptions,
and that close agreement between absorption spectrum and action
spectrum is to be regarded only as the ideal case. Actually the agree-
ment is sometimes very good as is illustrated by the action spectrum

for scotopic vision which, as shown in Figure 5, agrees nicely with the absorption spectrum of visual purple when proper account is taken of absorption by nonparticipating substances. In most cases, however, there is departure from such agreement, although the action spectrum reflects the character of the absorption spectrum to a greater or less extent. Obviously, all the assumptions employed in the analysis cannot always apply. For example, the light absorber probably absorbs much more than 10% in many instances—hence thin layers should be used when feasible—and, even if all the assumptions are fulfilled, there are still numerous other sources of error, some physical, some biological. The same conditions applying to rigid layers may apply to inadequately stirred suspensions of cells, leading to errors in the calculation of photochemical efficiency and in the interpretation of action spectra.

Numerous reasons have been mentioned above for expecting the absorption spectrum of a substance when measured outside the living system not to agree accurately with the absorption spectrum of that substance when inside the living system. There is also the possibility that a part of the incident light is absorbed by nonparticipating substances that absorb selectively with respect to wavelength. A good example of this is found in the erythema of sunburn, where the photochemical reaction takes place in the Malpighian layer of the epidermis beneath the horny layer (corneum), which strongly absorbs some of the wavelengths that elicit the erythema (3). Absorption by the corneum shows a strong maximum at 280 mμ, which no doubt accounts for the sharp minimum in the action spectrum at this point (see Fig. 3). A similar case is that of the action of ultraviolet radiation on pollen grains, where the pollen wall absorbs strongly the wavelengths that affect the living protoplasm within (25). Still another example is the absorption of shorter wavelengths by the lens of the human eye, which is illustrated in Figure 5. It is also possible for a nonparticipating substance inside the cell, in close association with the light absorber, to act as an internal filter; a still more difficult effect to estimate quantitatively.

It is obvious that the scattering of light by structures in the living organism may affect the absorption spectrum, since the proportion of the light scattered may vary greatly with the wavelength. Fluorescence is another possible source of difficulty, since not only does it serve to diffuse the light but introduces chromatic impurity, as the fluoresced light may have a considerably different wavelength from

that of the incident.   These factors, particularly the latter, are prob-
ably small in many cases, but they are possible sources of error in
some.

For any photochemical process there must be a limiting size of
quantum that supplies the necessary energy of activation or—in the
case of the endergonic photosynthetic process—the energy of reaction
as well.   This may limit effectiveness at the long wavelength end of
the spectrum, causing failure of agreement between action and ab-
sorption spectra.   It has already been pointed out that no photochem-
ical processes are to be expected in the pure rotation–vibration region
of the infrared.   In the case of inactivation of enzymes by ultraviolet
radiation, the quantum yield is not independent of the wavelength.
This may also be true in other cases, resulting in discrepancies be-
tween action and absorption spectra.

In addition to the "physical" sources of discrepancy just de-
scribed, there are "biological" sources.   One finds frequently, par-
ticularly in the older literature, descriptions of the "antagonistic"
effect of radiation from different regions of the spectrum.   While it is
possible that in a simple nonliving system certain wavelengths may
have a modifying action on that of other wavelengths—since different
parts of the light absorber molecule may be involved in different reac-
tions—it is probable that most if not all cases of "antagonism" in
living systems result from the complexity of these systems.   We may
cite the case of the germination of lettuce seeds studied by Flint and
McAlister (2, Ch. 3), in which the over-all effect was found to be
resolvable into three separate action spectra.   A somewhat different
example is that of inhibition of the erythema of sunburn by massive
doses of ultraviolet radiation.   The erythema is the manifestation of
the dilation of minute vessels in the dermis, in response to a dilator
substance elaborated in the more superficial Malpighian layer as a
result of the action of radiation of wavelengths shorter than about
$320 \, m\mu$.   The Malpighian layer itself contains no vessels; the shorter
of the erythema-producing wavelengths fail virtually to pass through
this layer, but a certain portion of the longer wavelengths penetrate
more deeply and reach the vessels of the dermis, which they appar-
ently inhibit from responding to the dilator substance (13,14).

Another instance of participation of more than one light absorber
is that in the blue-green alga *Chroococcus*, where Emerson and Lewis
(16) found that in addition to chlorophyll the pigment phycocyanine
acts as a light absorber in photosynthesis.

Color vision offers problems of considerable difficulty in studying the spectral vision of animals lower than man; failure to understand these seems often to have led to untenable conclusions. In photopic vision (cones) in man, one finds that, above the threshold, the matching of different hues is a function of factors other than the magnitude of the exciting radiation. For example, if two monochromatic beams from different parts of the spectrum fall upon the same area of the retina, the color sensation evoked differs from that evoked by either beam alone and the threshold of perception for this color may be quite different from those for the separate beams. As a result it is not possible accurately to determine an action spectrum for human photopic vision by matching monochromatic beams unless the wavelength differences are very small. In animals other than man, there is no direct way of knowing whether colors are perceived as such or not. Thus, if one attempted to determine the action spectrum by matching the effect of monochromatic beams in eliciting some end point, he might expect to obtain in the case of animals that detect colors an action spectrum that did not accurately reflect the absorption spectrum of the light absorber. In fact, he might anticipate that quite different action spectra would be obtained when different methods of measurement are used. This seems to be the case with regard to the honey bee, about which there exist two different sets of determinations giving spectral sensitivity curves differing almost completely except for the spectral range (11,23). A clue to the discrepancy appears when one examines the biological criteria used; in the one case the measurements consisted in the differential determination of the bee's response to two monochromatic beams; in the other, the response to a monochromatic beam was measured in a comparable manner against a polychromatic beam that was not spectrally homogeneous (black body radiator).

When one reviews the many factors that may affect the action spectrum and may interfere with an exact interpretation of its meaning, he may be inclined to wonder at the value of such determinations, for the few instances of close agreement between action and absorption spectra that may be cited are certainly the exception rather than the rule. But, nevertheless, such measurements constitute a necessary part of the study of any photobiological mechanism. The question is not whether attempts to measure action spectra should be made, but how carefully it is profitable to make them in a given case. Sometimes the nature of the system under study is such

that careful measurement of the effectiveness of incident monochromatic beams of high purity cannot yield direct information unless many factors that are not feasible to measure are taken into account; whereas in other cases, exact measurements may yield valuable information more directly. Thus, enthusiasm for very exact physical measurements should be tempered according to the problem in hand, particularly when the desire for perfection may inhibit obtaining less exact but more feasible measurements.

Some, but not necessarily all of the above mentioned errors, may be avoided when dealing with photochemical reactions in extracts from living organisms. Many useful applications of action spectra have involved such systems, e.g., the identification of enzymes by Warburg and others.

It must be admitted that in some instances very rough determinations of the limits of the action spectrum may be almost as valuable as more exact ones. These can often be obtained with cut-off filters, by simply determining the effectiveness of an appropriate source through short wavelength and long wavelength cut-off filters. Choice of filters and source depends upon the part of the spectrum concerned, and the intensities required. When very high intensities are needed to elicit a biological response, and also when it is necessary to expose more than small areas, the use of monochromatic radiation may not be feasible, and in such cases determinations with filters are particularly useful. When the incident radiation consists of a series of discrete lines, as in the case of the ubiquitous mercury arc, filters may yield as accurate estimates of action spectrum limits as a monochromator, in some cases possibly more accurate because they permit greater intensities to be used.

When a substance of known absorption spectrum is suspected of being the light absorber, an approximate test may be made by estimating the relative effectiveness to be expected when different filters are used. For a given wavelength, λ:

$$S = I_s T \epsilon \lambda \tag{22}$$

where S is the relative effective intensity, I_s is the relative intensity of the source, and T the transmission of the filter. The molecular extinction coefficient might be replaced by k or k'. The relative effectiveness S_r through a given filter transmitting wavelengths λ_1 to λ_2 is:

$$S_{\lambda} = \int_{\lambda_1}^{\lambda_2} I_s T \epsilon \lambda \, d\lambda \tag{23}$$

Values of S_x can be obtained graphically from curves such as those shown in Figure 7. This is a useful method where experiments have to be done out of doors with sunlight as the source, for example, when
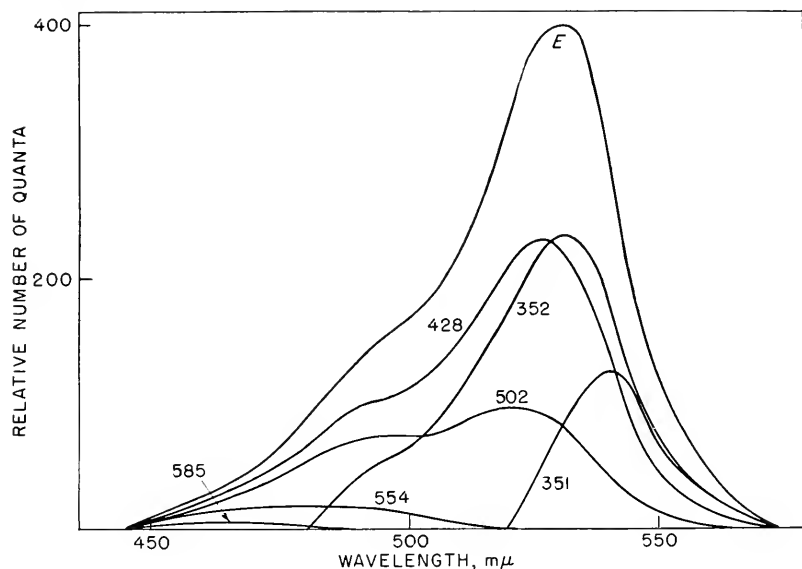
Fig. 7. Absorption (E) of radiation from a tungsten filament lamp of known spectral characteristics, by a solution of eosin ($I_s \epsilon \lambda$). Numbered curves are the same through filters of known transmission ($I_s T \epsilon \lambda$) (see equations 22 and 23) (*12b*).

photosensitization of domestic animals by plant pigments is in question (*2*, Ch. 12–15, 19; *27*). In this case data for the solar spectrum compiled by Moon (*21*) are useful.

1. General Remarks

The interpretation of an action spectrum may depend in considerable part upon biological information that does not enter into the measurements themselves. There are certain effects that may be grouped together, and there is a limited number of photochemically active substances to be found in biological systems; both facts must be taken into consideration when interpreting the experimental find-

ings. For example, wavelengths shorter than about 320 mµ may
cause damage to any living material. This is often attributed to
the fact that these wavelengths are absorbed by, and cause changes
in, two extremely important components of living systems, proteins
and nucleic acids. Action spectra for these effects have been very
carefully measured in numerous instances, yet they often leave much
to be desired as regards interpretation because of factors that are
unknown or cannot be taken quantitatively into account. In some
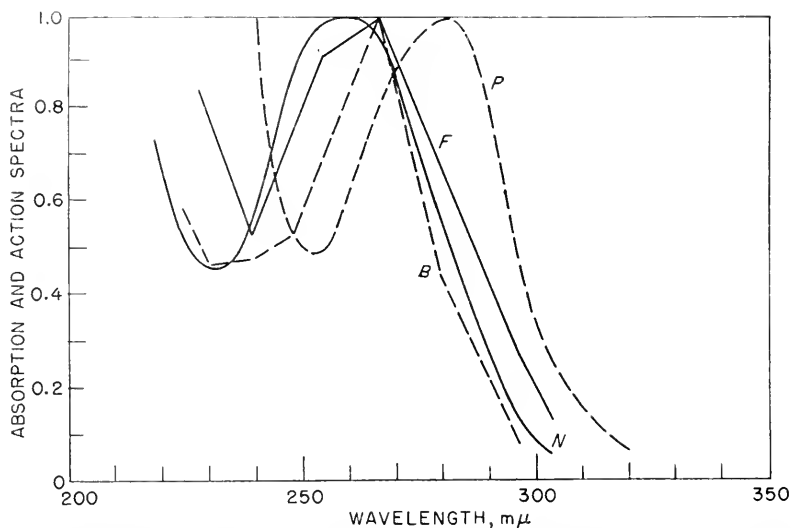
Fig. 8. Action spectra for killing of a bacterium (*B*) and a fungus (*F*) com-
pared to absorption spectra of nucleic acid (*N*) and of a protein (*P*). Ordinates
(units chosen arbitrarily to bring the maxima to a common value) for action
spectra are $1/I_0$ and for absorption spectra log (I/I_0). Original references are
cited in (*12a*).

cases there is good agreement with the absorption spectra of the
above-mentioned substances (*e. g.*, Fig. 8); but unfortunately for
the clarity of the picture, there are a great many substances that
absorb in this same region, *e.g.*, almost any compound containing a
phenolic ring. Moreover, precursors of the antirachitic vitamins (D)
absorb here and are activated by the same wavelengths, although the
action spectrum for antirachitic action shows characteristic differences,
consistent with the differences in the absorption spectra of these sub-
stances, from those of proteins and nucleic acids (compare Figs. 8

and 9). Thus one cannot say definitely on the basis of action spectra alone that the universally destructive effects of wavelengths shorter than 320 mμ are due to the action of these wavelengths on proteins and/or nucleic acids, although this is probably true in some cases.

Mutations are also produced by the same wavelengths, and there is reason to believe that the light absorber in this case is nucleoprotein, as action spectra suggest (19), because of the close association of these substances with the genetic mechanism of the cell.

When photosensory responses are concerned, whether phototropic bending of plants or visual mechanisms in animals, one may on various
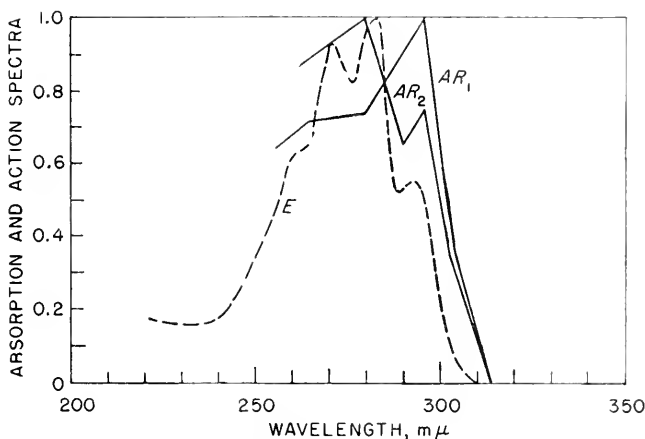
Fig. 9. AR_1 and AR_2 are action spectra for antirachitic action in the rat (determinations of two different sets of investigators). E is the absorption spectrum of 7-dehydrocholesterol in ethyl alcohol. Ordinates (units arbitrarily chosen to bring maxima in agreement) for action spectra are $1-I_0$ and for absorption spectra, $\log (I/I_0)$. Original references are cited in (12a).

grounds, including evolutionary reasoning, suspect carotinoid pigments (26). Tetrapyrrole ring compounds, including chlorophyll and porphyrins, are widely distributed in living organisms and likely to be photoactive, but are by no means universally so, nor are the porphyrins the principal offenders in human photopathology they were once thought to be (2, Ch. 19).

Examples could be multiplied, but it is best perhaps to close instead with a brief word of caution—that in this field, as elsewhere in biophysics, exact physical measurements are apt to be of limited usefulness unless linked with sound biological interpretation.

References

GENERAL REFERENCES AND REVIEWS

1. Bhagavantam, S., *Scattering of Light and the Raman Effect*. Andhra University, Waltair, 1940.

2. Blum, H. F., *Photodynamic Action and Diseases Caused by Light*. Reinhold, New York, 1941.

3. Blum, H. F., *Physiol. Revs.*, **25**, 483–530 (1945).

4. Bowen, E. J., *The Chemical Aspects of Light*. Clarendon Press, Oxford, 2nd ed., 1946.

5. Brode, W. R., *Chemical Spectroscopy*, 2nd ed. Wiley, New York, 1943.

6. Caspersson, T., *Skand. Arch. Physiol.*, **73**, Supplement No. 8 (1936).

7. Lewis, G. N., and M. Calvin, *Chem. Revs.*, **25**, 273–328 (1939).

8. Miller, E. S., *Quantitative Biological Spectroscopy*, Vol. I, *Absorption Spectra*, 2nd ed. Burgess, Minneapolis, 1940.

9. Rollefson, G. K., and M. Burton, *Photochemistry and the Mechanism of Chemical Reactions*. Prentice-Hall, New York, 1939.

9a. West, W., "Spectroscopy and Spectrophotometry," in *Physical Methods of Organic Chemistry*, 2nd ed., A. Weissberger, ed. Interscience, New York, 1949, Chapt. XXI.

10. Wood, R. W., *Physical Optics*, 3rd ed. Macmillan, New York, 1934.

SPECIFIC SUBJECTS

11. Bertholf, L. M., *Z. vergleich. Physiol.*, **18**, 32–64 (1932).

12. Blum, H. F., *J. Natl. Cancer Inst.*, **3**, 539–543 (1943).

12a. Blum, H. J., in O. Glasser, *Medical Physics*. Year Book Publishers, Chicago, 1944, pp. 1145–1157 (2nd ed. in press).

12b. Blum, H. F., and K. G. Scott, *Plant Physiol.*, **8**, 525–536 (1933).

13. Blum, H. F., and W. S. Terus, *Am. J. Physiol.*, **146**, 97–106 (1946).

14. Blum, H. F., and W. S. Terus, *Am. J. Physiol.*, **146**, 107–117 (1946).

15. Emerson, R., and C. M. Lewis, *Am. J. Botany*, **26**, 808–822 (1939).

16. Emerson, R., and C. M. Lewis, *J. Gen. Physiol.*, **25**, 579–595 (1942).

17. Emerson, R., and C. M. Lewis, *Am. J. Botany*, **30**, 165–178 (1943).

18. Hollaender, A., and W. D. Claus, *J. Gen. Physiol.*, **19**, 753–765 (1936).

19. Hollaender, A., and C. W. Emmons, *Cold Spring Harbor Symposia Quant. Biol.*, **9**, 179–186 (1941).

20. Kirby-Smith, J. S., H. F. Blum, and H. G. Grady, *J. Natl. Cancer Inst.*, **2**, 403–412 (1942).

20a. Miller, E. S., *Plant Physiol.*, **12**, 667 (1937).

21. Moon, P., *J. Franklin Inst.*, **230**, 583–617 (1940).

22. Rentschler, H. C., R. Nagy, and G. Mouromseff, *J. Bact.*, **41**, 745–774 (1941).

23. Sander, W., *Z. vergleich. Physiol.*, **20**, 267–286 (1933).

24. Smakula, A., *Angew. Chem.*, **47,** 657-665 (1934).

25. Uber, F. M., *Am. J. Botany*, **26,** 799-807 (1939).

26. Wald, G., *Harvey Lectures*, **61,** 117-160 (1945-46).

27. Riemerschmid, G., *Onderstepoort J. Vet. Sci. Animal Ind.*, **17,** 89-104 (1941).

28. Wald, G., *Science*, **101,** 653-658 (1945).

29. Wald, G., in Proceedings of the Colour Vision Conference, Cambridge, England, 1947; *Documenta Ophthalmologica*, Vol. 3 (in press).

Addendum

Since this chapter went to press there have been two important developments which should be mentioned briefly:

Reinvestigation of the quantum efficiency of photosynthesis by Burk, Hendricks, Korzenovsky, Schoken, and Warburg (*Science*, **110,** 225-229 (1949)) has yielded maximum values near 4 quanta per molecule of oxygen liberated, in good agreement with the values reported by Warburg in 1923. This calls for some modification of values cited in this chapter.

Enhancement of recovery from effects of ultraviolet radiation by exposure to light from the short end of the visible and long end of the ultraviolet spectra has been demonstrated: in fungi by Kelner (*Proc. Natl. Acad. Sci. U. S.*, **35,** 73 (1949)); in bacteriophage (in the presence of *E. coli*) by Dulbecco (*Nature*, **163,** 949 (1949)); and for the sea urchin by Blum, Price, Robinson, and Loos (*Biol. Bull.*, **97,** 232 (1949)), and Marshak (*Biol. Bull.*, **97,** 244 (1949)). This may call for revision of our ideas on the effects of ultraviolet radiation.

X RAYS AND X IRRADIATION

JOHN W. GOWEN, *Iowa State College*

Three broad categories of investigators are interested in the biological effects of X rays. First are those who wish to photograph or register on fluorescent screens abnormal structures deep within the tissues. In this manner abnormalities are brought to view and suitable action taken. The second group is interested in the effect of radiant energy on abnormal growths within the organisms. It seemed possible that, through differential absorption of radiant energy, abnormal growths could be immobilized, and eventually regress and be replaced by normal tissue. Some success in this direction has led to a broad study of X-ray effects. The third group is

interested in the effect of radiant energy on normal cells and cell structures. These studies first investigated the effects of radiant energy on the cell lineage pattern in the unfolding of embryological developments. The studies were confined to particular tissue cells, as for instance the sperm and egg, nucleus, or cytoplasm. The microscopic changes observed in these tissues as the exposure progressed were used to measure the radiation effects. Different effects in particular tissues pointed to a selective action of the X rays. The need for understanding radiation effects in the regeneration of particular organs became evident and such investigations soon paralleled those on embryological development. Somewhat later, geneticists became interested in the more intimate phases of X-ray effects within individual cell nuclei. X rays furnished the means for penetrating into the nucleus and exciting particular elements, with the consequence that genes and chromosomes were changed and the course of heredity modified.

The three types of investigators want very different results from the radiant energy. The first, by using X rays simply as a means of differential diagnosis of abnormal structures, wish X rays to penetrate without producing any obvious effects on the tissues under observation. All that is required is an accurate, contrasting photograph of the normal and abnormal structures. The second group of investigators, on the other hand, require the X rays to penetrate to the abnormal tissues and be absorbed in them to a sufficient extent to cause their death and regression without affecting the normal surrounding tissues. Finally, those who wish to follow the embryological unfolding of organisms through changes produced by radiation, or those in the genetic field who wish to alter genes or chromosomes require the X rays to penetrate to the cells under observation, cause changes in these cells that will lead to the sought-for structural changes in the adult, and yet leave the cells in sufficiently normal condition so that they will go on and reproduce, either to form the complete new body phenotype or to be carried by the germ cells to a following generation, when the altered genes would produce a distinguishable change.

Besides these main biological interests, X rays have been used for the elucidation of molecular structure. X-ray diffraction photographs of crystals have given information on the arrangement of the atoms within molecules. Similarly, in larger molecules it is possible to estimate the geometrical shape and size of molecules from crystal

or powder X-ray patterns. To the biologist this information has been most important in the study of viruses or biologics, as penicillin, when crystals may be formed, or in the analysis of elastic fibers.

Summary volumes covering some of these points are listed in the general bibliography (1–4).

A. SOURCES OF X-RAY RADIATION

It is presumed that the readers of this volume will be interested primarily in investigating the biological effects of X irradiation rather than in the construction and maintenance of X-ray apparatus. Many will depend on hospitals or physical laboratories, where regular X-ray equipment is being maintained, for their source of radiation. Consequently, only a minimum acquaintance with X-ray tubes will be needed. Where possible, the irradiations had best be made in a physical laboratory. The biological material will then have the benefit of trained specialists in both the biological and irradiation fields. The necessary conditions of tube operation will be assured and proper measurement will be made of the dose of X rays applied to the tissues. But, since investigators are not equally fortunate, it is well for all to be aware of some of the necessities of X-ray exposure and protection. A few major points may be mentioned.

The early X-ray tubes were gas-filled types operating at relatively low voltages. These tubes delivered radiant energy of relatively long wavelengths and, therefore, had low penetrating power. The biological effects were largely surface effects. With further development and water cooling these gas-filled tubes became highly effective generators of X rays in the wavelength region from 8 A. (an angstrom unit = 10^{-8} cm.) to 0.3 A. Owing to their long wavelength, the X-radiation dosage rate in ionization units (roentgen or r. unit, defined later) delivered by these tubes can reach very high values.

The constructional features of a gas X-ray tube are reproduced in Figure 1. For contrast, Figure 2 shows a vacuum type tube of a closely similar design. The cathode (C) focuses the high speed electrons on the target or anode, thus generating the X rays, which are emitted from the window at K (47). The low gas pressure in this tube, about 10^{-4} mm. of mercury, is controlled by vacuum pumps and an adjustable air leak. The water-cooled anode may be composed of any suitable metal desired. An anode metal is often

selected for the intense characteristic radiation that it emits in a
limited wavelength region, *e.g.*, a copper anode has characteristic
wavelengths near 1.5 A. Gas-filled X-ray tubes can be made self-
rectifying, thus requiring only a transformer in their electrical cir-
cuit, but the difficulty of maintaining satisfactory air pressure makes
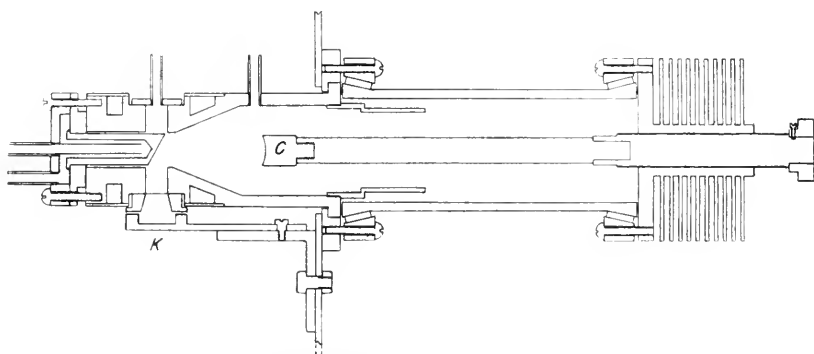
Fig. 1.   Self-rectifying gas type X-ray tube. (Courtesy L. E. Pinney, *47*).
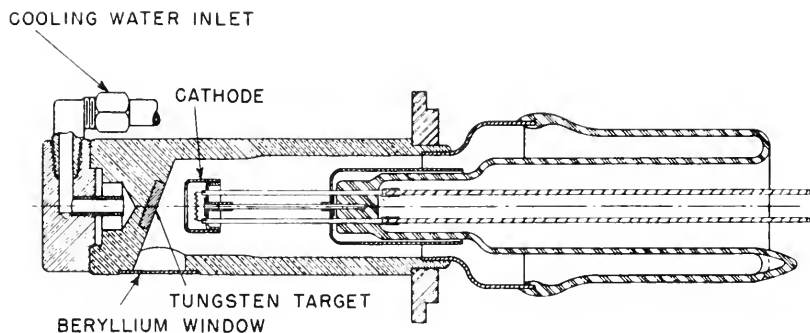
Fig. 2.   Filament type X-ray tube, evacuated and sealed (*56*). (Courtesy
Machlett Laboratories, Inc.)

gas tubes undesirable for occasional laboratory work except where
very large dosage rates are essential to an experiment. Gas tubes
have a further limitation in that they will not operate at voltages
higher than 50-60 kilovolts. This restricts their use in experimental
work since the operating voltage determines the short wavelength
limit of the X rays emitted.

Coolidge type X-ray tubes (see Fig. 2) with sealed vacuum have
almost entirely replaced gas-filled tubes for most laboratory work as

well as for diagnostic and therapeutic uses. Their advantages are better control, lower maintenance, and constant readiness for use. The thermionic emission of electrons from the cathode, which consists of a filament heated to incandescence by an electric current, permits convenient control of the X-ray output. The electrons are focused by the cathode to strike the anode at a small spot. The X rays pass out through the thin wall of the tube or through a special window. The Coolidge type tubes are now available in a wide range of designs and power ratings. Water-cooled models are preferred for heavy duty continuous operation. They permit higher dosage rates in that the samples can be placed closer to the anode, thus avoiding much loss in intensity from the inverse square law factor. The form of the radiation intensity versus wavelength curve depends on several factors besides the applied voltage to the tube. These will be discussed later. Ordinarily, voltages are confined within the range of 30 to 250 kilovolts. Lower voltage tubes with especially thin windows are also available for special purposes.

For higher voltages special type equipment is available in some laboratories and hospitals, as, for instance, the Sloan tube and generator. These tubes are capable of delivering 850 to 1000 kilovolt X rays. The radiation is general, covering all wavelengths up to the limit of the exciting voltage. The Sloan apparatus (5, p. 85) involves an oscillator circuit with auxiliary tubes similar to those of a short wave broadcasting set. For still other types of tubes, see Clark (1).

Betatrons and synchrotrons are two of the newer types of equipment from which it is possible to obtain X-ray radiation of 10–100 million volts. Their use in biological experiments is just beginning. Current physics periodicals must be consulted for details. Their widespread use is expected, at least for industrial purposes.

When X radiation is obtained from a qualified physical laboratory, there probably will be no difficulty in the operation of the apparatus or in obtaining the energy output. For the investigator who wishes to do experimental work using hospital equipment the difficulties may be greater. In general, he must acquaint himself with the type of X-ray machine and its suitability for the purpose he has in mind. One point to be carefully taken into consideration is that for the treatment of biological materials the machine should be capable of continuous operation over the necessary time without damage to the tube or its power supply. This requirement precludes

the use of such X-ray equipment as is often found in dental clinics, where it is intended for intermittent operation only, with separate exposures limited to a matter of seconds.

The protection of the operator from the X rays generated by the X-ray machine is of first importance. Lead shields are normally provided for this purpose. It is well to read the bulletins of the Bureau of Standards on suitable protection of the operator from X-ray effects (55).

B. DOSIMETRY

1. Units of Measurement

Since the discovery of X rays, the measurement of their intensity has been conducted along two quite different lines. For biological measurements the unit has changed as increasing information became available on the biological effects of this energy. One of the first noticed effects of X rays was an erythema or a reddening of the skin, occurring either immediately or as a delayed reaction appearing after some days or weeks. To the clinician this effect was obviously the beginning of cell changes that were rather serious to the tissues. The "erythema dose," defined as the minimum X radiation to produce an erythema after ten to fourteen days, became a biological measure of X-ray dosage. It suffered the disadvantages that people varied in their susceptibility to radiation and that the unit differed according to whether radiation was delivered to the skin continuously or intermittently.

A second biological unit used for X-ray measurements is the survival of such objects as *Drosophila* eggs. These survival tests were found to be remarkably constant in eggs from rapidly laying *Drosophila* exposed two hours after laying. The unit was the X radiation necessary to cause the death of 50% of the eggs exposed under standard conditions (43). The biological measures of X radiation are handicapped by lack of trained personnel and easily available material, as well as from the time elapsed between the radiation exposure and a knowledge of the dose administered.

The physical measurement of X radiation, on the other hand, can be made in a very few minutes with easily standardized equipment. Until the biological and physical methods were brought into agreement, however, no common ground for the use of one method over the

other existed.   This gap has been filled through the work on *Droso-
phila* eggs (*43,44*) followed by that of genetics workers on gene muta-
tion and on survival of *Drosophila* sperm (*26*).   The measurements
are in agreement and show that the total energy absorbed in the tis-
sues as the X rays pass through them is the important element in
producing the effects.   The physical unit of X radiation, r. unit or
roentgen, is the quantity of X rays "such that the associated corpus-
cular emission per 0.001293 gram of air produces, in air, ions carrying
1 electrostatic unit of quantity of electricity of either sign."   The
dose in r. units, therefore, is not a measure of the total energy in an
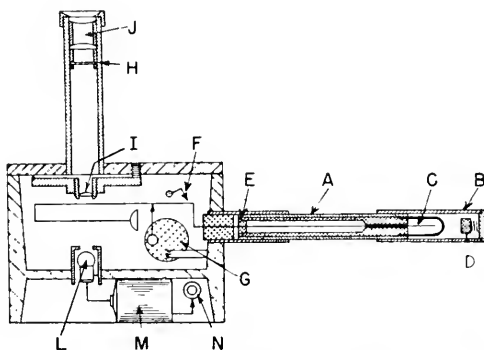
Fig. 3.   Sectional view of an X-ray dosimeter or r. meter (courtesy
Victoreen Instrument Co.):   A, chamber tube containing condenser
and chamber;   B, chamber tube cap (for end opposite chamber);   C,
ionization chamber;   D, insulating button to cover insulation at E;
F, discharge switch;   G, static charger wheel;   H, scale;   I, objec-
tive;   J, ocular;   L, lamp;   M, battery;   N, light switch.

X-ray beam, but only that fraction which can be dissipated by the
ionization process in a thin layer of air.   This can be related to the
energy dissipated in tissue by ionization.   For a further discussion,
the reader should consult the following chapter.

A diagram of a small dosimeter based on this ionization principle
is shown in Figure 3.   The ionization chamber is a small thimble-
shaped structure at the end of the tube.   Its walls are easily pene-
trated by X rays, the charge being taken by the wire to the electro-
scope.   The current causes the electroscope to discharge;   the amount
of the discharge is measured on an r. unit scale, which is read by look-
ing through the microscope.   This type of dosimeter must be cali-
brated for the quality of X rays absorbed, however, by comparison

with a standard ionization chamber. The whole box is heavily shielded against X rays entering at any place other than the thimble chamber.

The threshold erythemal dose is equivalent to about 500 r. since this dose delivered at one sitting to a local skin area is capable of producing perceptible reddening in 80% of patients within four weeks of treatment. Similarly, the dose necessary to kill 50% of the *Drosophila* eggs under the standard condition is 180 r.

2. The Depth–Dose Problem

The close relation between the biological effects and the X-ray dose in r. values in quantitative experimentation justifies the measurement of X-radiation doses in r. units. For actual dose measurements in r. units to correspond to the absorbed energy in the experimental tissue, certain conditions must be satisfied. If the layer of material to be treated is thin, as, for instance, a film of virus on Cellophane, the experimental arrangement must be such that the film is treated with a sufficient layer of air between it and any other more dense material to prevent the return of scattered radiation from the dense substances to affect the virus.

This "back-scatter" effect becomes particularly important in treating deep tissues. Problems are presented not only by the scattering characteristics of the tissues that surround the object of investigation but also by the absorbing power of intervening layers of tissue. For example, if a bone is located close to deep tissue, scattered radiation from the bone would change the dose from that expected if the bone were not present. Computations of such an effect are very difficult because the X rays are not monochromatic and the several materials present of different atomic number contribute in various amounts to the back-scattered radiation. With an appropriately constructed "phantom" (see 5, p. 179), dosage measurements to take such effects into account are not too difficult to make. At the same time it can correct for the layer of material over the treated object, which is likewise important in order to find the dose incident on the object under investigation. The total "absorption" in this layer depends on several physical phenomena, including photoelectric effects and two types of scattering. These different phenomena affect the absorbing power of a substance with respect to its atomic number in different ways. The photoelectric absorption, for example,

varies approximately as the fourth power of this number. Since the
atomic number of lead is high, a comparatively thin lead screen may
give effective protection against X rays (55). In tissue, the absorp-
tion can be represented by a weighted averaging of the atomic num-
bers of the atoms composing it.

Another complicating factor in this absorption correction is that
of the wavelengths of the X rays. It is not practical to obtain high
intensity monochromatic radiation with X rays for irradiating the
large areas encountered in most biological work. X-ray tubes give
out a continuous band of general radiation having wavelengths rang-
ing from 10 A. to some limiting value dependent upon the voltage put

Fig. 4. Variation of intensity
of general X radiation with
wavelength. Characteristic radi-
ation is superimposed. Dashed
line shows intensity distribution
possible by using proper metal
filter.
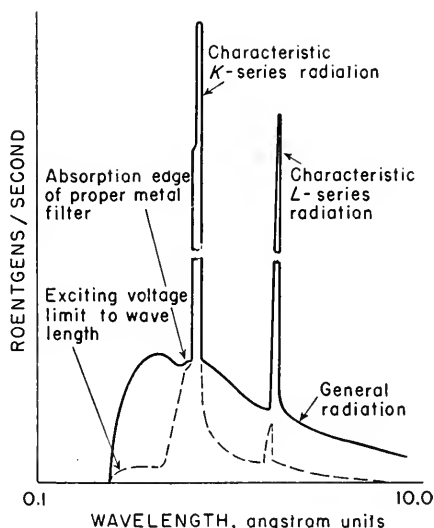
on the tube. A graph showing the absolute intensity as a function
of wavelength for general radiation is shown in Figure 4. General
radiation shows a continuous spectrum of wavelengths from a low
intensity value in the long wavelengths, rising to a modal value, and
then dropping more sharply to zero intensity at the short wavelength
cut off. This limiting wavelength in A. is given by $12,395/V$, where
V is the voltage. With 20 kv. on the tube, the wavelength limit
would be about 0.6 A.; with 200 kv., 0.06 A.

Besides general radiation there is superimposed radiation charac-
teristic of the metal used as target in the X-ray tube. Characteristic

radiation occurs in a series of peaks, K, L, M, etc., which correspond to the various electronic energy levels in the atom. For copper the K peak is at 1.5 A.; for silver, at 0.49 A.

Interposing a sheet of metal between the object and the X-ray tube has the effect of filtering out the radiations of the longer wavelengths, but it also results in the emission of secondary X rays which are characteristic of the atomic number of the metal used as a filter. These secondary X rays are always longer in wavelength than the original X-ray beam. In case they are absorbed by a filter of still lower atomic number, perhaps aluminum, characteristic X rays of aluminum are emitted, but of course are still longer in wavelength.

The significance of these facts to the investigator comes in emphasizing the experimental arrangement that must be made for critical experiments and the care with which the measurements of the X rays received by the specimen must be taken.

It is not always possible to meet these conditions exactly. An investigator interested in irradiating *Drosophila* sperm and then utilizing them for fertilizing *Drosophila* eggs requires sperm irradiated in the intact adult male fly. X radiation incident to the fly can be estimated in r. units, but absorption in the chitinous body wall that surrounds the abdomen of the fly and back-scatter of the tissues to the sperm become difficult to measure. The dosage measurements could be made from two positions, *i.e.*, at the body surface before any X radiation has passed through the fly's tissue and after the radiation has passed through the fly's body. The difference between these measurements shows the amount of radiation absorbed in the fly's tissues. Neither of these measurements is what the investigator wants. He wishes to know the radiation incident to the sperm. Two assumptions might be made to facilitate this estimate: that the sperm are in some definite position within the body (as at its center) and that the layers of tissue above and below the two surfaces are absorbing uniformly and with negligible back-scatter. X-radiation absorption proceeds exponentially as follows:

$$I = I_0 e^{-\mu x}$$

where I_0 is the intensity of the radiation incident to the fly, and I is the intensity that passes through the fly; x is the thickness of the fly and μ is an "average" absorption coefficient of the fly. Then $I_c = I_0 e^{-\mu x/2} = I_0 (I/I_0)^{1/2}$, where I_c is the intensity at the center of the fly.

For several different flies the I/I_0 value for silver X radiation was 0.894; for copper, 0.530; and for chromium, 0.246. The square roots of these values give 94, 72, and 49%, respectively, of the incident surface radiation reaching the center of the fly. The respective X radiations from these three metal targets were filtered through palladium, nickel, and vanadium windows. The dose incident to the fly must be corrected by these percentages to get the estimated dose to the sperm.

3. Monochromatic X Radiation

The difficulties attached to studies of wavelength effects bring out other points for consideration. For these studies one would like monochromatic X radiation. Unlike visible light, an X-ray beam cannot be readily separated into its constituent wavelengths. Monochromatic X radiation thus far has had to be obtained by diffraction from crystals or gratings. This results not only in a great reduction in intensity at each wavelength but also in a very small area over which the X rays are monochromatic. These facts make this method impractical for most biological work. As the next best alternative the experimenter attempts to restrict the radiation to particular bands. This is accomplished largely through filtering out the X rays of long wavelength and through the use of particular metals as windows. The filters for the long wavelength components are generally made of aluminum or copper. The general radiation reaching the tissues is therefore restricted in some degree to the short waves of the spectrum.

Besides this general absorption effect, there is a characteristic absorption effect. Every element has a characteristic absorption curve; these curves are marked by discrete peaks or discontinuities at specific wavelengths. Use is made of this fact in the choice of filter materials for X-ray windows. In using a copper target tube, for example, the continuous background of general X radiation has imposed upon it radiation characteristic of copper. The "K"-characteristic radiation of copper occurs at a wavelength of approximately 1.5 A. The presence of this characteristic radiation has the effect of making a large peak in the radiation curve when intensity is plotted against wavelength (cf. Fig. 4). A thin sheet of nickel employed as a window in the X-ray tube thus possesses the unique property of transmitting nearly all of this copper peak radiation at 1.5 A.

while filtering out much of the general radiation on either side of this peak. Consequently, the radiation after this filtration is more nearly monochromatic. The actual choice of a filter depends largely on its atomic number. In general, the element of choice is one or two atomic numbers below the metal used as the anode of the X-ray tube. Palladium is used to filter X rays from silver anode tubes, zirconium for molybdenum, nickel for copper, manganese for iron, and vanadium for chromium. The short wavelength limit of this method of filtration occurs at about 0.1 A.; it is applicable, then, only to X rays of medium or long wavelength. Graphical representation of the effects is shown in Figure 4.

A further difficulty in obtaining monochromatic X rays is the alternating potential ordinarily used to operate the tubes. The alternating rise and fall of the high voltage supply (cf. 5, Fig. 57) results in a consequent shortening and lengthening of the wavelengths of the X rays produced. Various schemes are used to reduce this effect, but some voltage ripple is almost always present. If adequate rectification is not available to eliminate this effect, then, at best, filter techniques will lead to a rather broad band of wavelengths. Direct current sources could be used, of course, especially for low voltage studies.

For some biological investigations, the Ross (4, p. 532) double filter technique might prove useful. It is based on differential absorption by metallic filters.

C. THEORIES OF X-RAY ACTIONS

1. The "Point-Heat" Concept

Historically, the first general theory designed to explain the results observed in the treatment of tissues by X rays was that of Dessauer (15). It was recognized that radiant energy, in passing through gases or tissues, was subdivided into discrete units as it underwent absorption in these substances. A particulate or quantum theory of X-ray absorption was evolved. At the point of release of the energy of these quanta, two types of effects were conceived. The first looked on the absorbed energy as generating a considerable amount of heat at the small focus of the release. This theory was developed into the "point-heat" theory of X-ray effects by Dessauer. The theory accounted for the very specific and localized effects of the X

rays on the tissues. The survival curves, observed when bacteria
or other forms were irradiated, could be explained by a rationalization
of the Poisson equation (see Sect. G). The X rays were visualized
as distributed over the bacteria or, in general, the target as discrete
particles and in a random manner. Absorption of these X-ray par-
ticles was regarded as causing point-heat lesions in the tissues. For
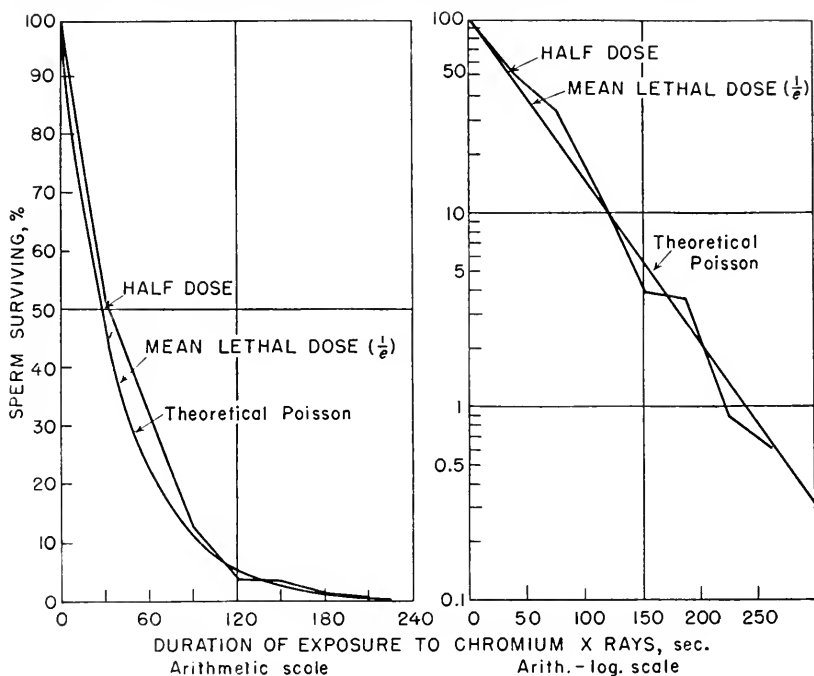
Fig. 5. Effect of irradiating *Drosophila* sperm with X rays from chromium
target tube: at left, survival data are plotted on linear percentage scale with doses
of X rays in r.; at right, the ordinate is logarithm of survival. Actual data curve
is irregular line. Best fitting Poisson equation is smooth curve. Advantage of
arith.-log grid at right is immediately evident in that Poisson equation on this
grid gives straight line plot.

a given organism the chance that one of these absorptions would
fall in a spot vital to the organism was constant for the same amount
of X rays applied under the same conditions. Similarly, this chance
was identical with that of another bacterium of the same species
and under the same conditions. For a unit dose of X rays, the chance
of death would be the percentage receiving these point-heat absorp-

tions or "hits." For a second unit of X rays the percentage receiving the point-heat absorptions would be identical with that of the first unit. However, this percentage would be distributed over organisms that had already received one such point-heat absorption and therefore were already incapable of surviving, and over a second group that up to that time had escaped such hits. The observed percentage of deaths would be less for the second unit than for the first unit of X rays. For a third unit of X rays on the treated organisms there would again be the same percentage number receiving a point-heat absorption as for the initial or second group. Again, however, these point-heat absorptions would be distributed at random over the organisms that had received such absorptions from the first or second units of X rays and so could not show the effects of absorption from the third dose of X rays. Only organisms that had thus far escaped receiving a "hit" could affect the subsequent death rate of the bacteria. The number dying on the third dose would be, again, less than that of the second dose and still less than that of the first dose. This would lead to a series of survivals having the general form of Figure 5. These results may be expressed generally by the successive terms of the Poisson distributions, where we can calculate the probability for receiving any of these point-heat absorptions and therefore the chance that the bacteria survive any dosage. The general form of these curves for actual data is as shown in the figure.

The interpretation of the effects of X rays as terms in a Poisson series is frequently used today. The point-heat concept, however, has been radically modified.

2. The Sensitive Volume Concept

A second development came in the work of Crowther (10,11). He showed that a large number of X-ray quanta must be absorbed by a protozoan in order to produce death. This led to the concept that the protozoa were composed of volumes vital to their functioning and volumes that could be destroyed without seriously injuring the organisms. The size of these vital structures was much less than the size of the organism. The organism was visualized as having a sensitive spot within it. This concept did not modify the Poisson interpretation put on the survival curve. But the observed data did add to the complexity of the interpretation, for the protozoa did not die according to the simple one absorption type of Poisson series.   40

absorptions in the vital spot were assumed to produce death. These data required a Poisson distribution that allowed organisms to survive a large amount of X rays before any deaths were noted. When deaths started they increased rapidly and then later decreased at about the same rate after they had passed the midpoint. The form of the survival curve was explained by the necessity of multiple absorptions of radiation in the sensitive spot before any observed effects were noticed. Figure 7 shows this and other curves plotted on the arith.-log. grid.

Condon and Terrill (9) extended the theory by pointing out that the size of the X-ray quanta increases with decreasing wavelengths. The energy of the X-ray quanta incident to the tissue is inversely proportional to their wavelength; on an energy basis a given biological effect might require several quanta of long wavelength X rays in the sensitive spot, where but one quantum would be necessary with short wavelength X rays.

3. Current Ionization Concepts

Further developments came with more detailed examination of how X-ray energy is released within tissues and gases. The absorption of a quantum in tissue may generate a photoelectron that carries nearly the full energy of the quantum. This photoelectron progresses through the tissues, occasionally colliding with electrons of atoms in its path, until all of its excess kinetic energy is dissipated in the formation of positive and negative ion pairs. The track or path of the photoelectron is very crooked and irregular. The length of any electron's path is dependent upon its initial energy and on how much energy is lost in each collision. The ion pair density along an electron's path becomes greater as the residual energy of the electron decreases. With X rays generated at potentials to 10 kv., their absorption by matter gives rise to photoelectrons almost exclusively.

When the initial energy of the quantum becomes larger, say 50 e.kv. (electron kilovolts), the energy distribution takes a somewhat different form. Then about 68% of the energy is distributed as photoelectrons, the rest as recoil electrons (Compton effect, see 4). A recoil electron is formed when an X-ray quantum collides with an electron in an atom, causing the electron to fly off at an angle from the initial path of the quantum. Recoil electrons ionize in the same manner as the photoelectons, but on the average have much less en-

ergy. The average length of these recoil electron paths is short compared to that of a photoelectron from a quantum of the same original energy; a recoil electron, for instance, from a 50 e.kv. quantum has an average energy of only 11 e.kv. Recoil electrons are visualized as distributing the energy obtained from high voltage quanta to many discrete spots in the absorbing tissue, in a manner similar to that of photoelectrons with the same amount of energy. Quanta with 100 e.kv. energy have about 72% of their total energy dissipated as recoil electrons, the mean electron energy being about 17 e.kv. With 1000 e.kv. quanta, only recoil electrons are formed; their average energy
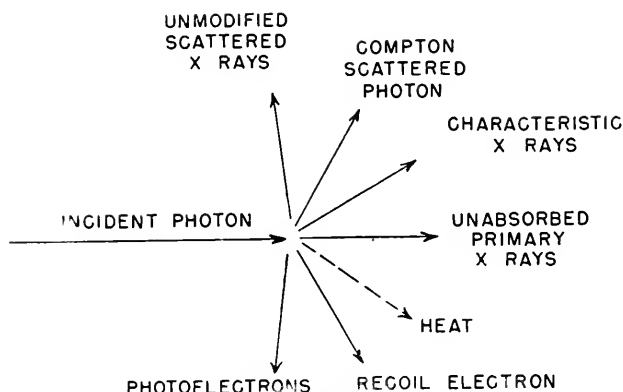
Fig. 6. Sketch to indicate multiple mechanisms of energy dissipation from an X-ray beam traversing tissue.

is about 435 e.kv. The recoil electrons, except for directional characteristics, behave essentially as photoelectrons in that their energy is dissipated by creating secondary electrons of still lower energy and by creating ion pairs along their path. The degradation scheme of X-ray energy is graphically represented in Figure 6.

The conversion of the energy of large quanta into recoil electrons rather than photoelectrons obviously could result in a very great difference in the observed effects of the X rays. For example, a theoretical basis for wavelength effects in biological investigations will tend to disappear. Work of Packard (43) on *Drosophila* eggs gave experimental support to this viewpoint since he was unable to detect any differential quantitative effect from the various wavelengths of X rays used.

Wyckoff (52) observed that colon bacteria were killed by X-ray

wavelengths lying between 4 and 0.5 A. Since semilogarithmic survival curves were obtained, he interpreted this to mean that the absorption of one energy unit in the vital volume of a bacterium was sufficient to kill it. On this assumption, only 1 quantum in 4 with the 0.5 A. X rays was necessary for death, whereas it took 60 quanta for one to be deadly with the 4 A. X rays. Instead of being more effective as a killing agent the longer wavelength X rays turned out to be somewhat less effective. On the basis of these and other experiments, the quantum does not appear to be the primary agent in the causation of the biological effects of radiation.

Another model for explaining the effects of radiation was furnished by Glocker (20). This model attempts to take into account the spatial distribution of the energy within the tissues as well as the quantum absorption. The photoelectron path through the tissues is an irregular one in which the ions per unit length of path are few in its initial stages, becoming greater as the energy of the electron drops, until at the end of the path the ionizations are close together. With this model a hit is considered to occur when the ionizing particle passes through the target (see also 3, p. 88; and 40).

More recently consideration has been given by Lea (3) to the fact that ion pairs are apt to form in clusters along the radiation track. Either one or several of the ions within a cluster may be considered sufficient to produce a given effect of irradiation. Different energy particles have differing clustering effects along their ionization paths. Clustering tends toward more than one ionization in a sensitive region if that region is large. It follows that, if the volume of the sensitive region is calculated on the basis of the total ion pairs formed, the size of the region is underestimated. If, on the other hand, the sensitive region is small, a single ion from a cluster may be sufficient to produce the observed results and at the same time lie outside the effective range of the other ion pairs in the cluster. This will lead to the correct size of the sensitive region. For small targets the method of Crowther (10), in which the ionizations are considered to be distributed at random, each ionization being capable of producing the observed effect, would lead to approximately correct results as far as size is concerned.

Where the biological material irradiated consists of several discrete parts, Lea has shown that it is possible to estimate the number of these parts as well as the volume of each from the X-ray results. This is accomplished by using irradiation with widely spaced ion

clusters, as, for instance, γ rays, as contrasted with densely spaced ion clusters found in α rays.   This method may be used to supplement other means of estimating the number of the sensitive volumes within an organism.   The estimate is subject to considerable error;   both square and cube terms are introduced into the calculations.

Another means of producing biological effects by X rays is that of raising the energy level of an electron in an atom or molecule.   In this unstable state chemical reactions may occur that would not occur otherwise.   With small molecules excitation is often as effective as ionization, but with large molecules excitation appears to be much less effective.   Biological material offers possibilities for both types of reactions.

X-ray energy differs from ultraviolet energy in the manner in which it is absorbed as well as in the type of effects produced.   Ultraviolet energy is absorbed selectively by certain cell constituents, whereas the X-ray energy is distributed at random along ionization tracks but in a more homogeneous manner over the tissue irradiated. The X-ray energy may produce either ionization or excitation, whereas the effect of ultraviolet probably is confined to excitation.

D. MATERIALS FOR STUDY OF IRRADIATION EFFECTS

Experience has shown that for the adequate interpretation of X-ray effects it is frequently necessary to follow the exposed cell and its progeny through many cell generations so that various delayed effects of the X rays may appear.   For critical quantitative experimentation the irradiated material should be composed of single self-reproducing entities.   With the wealth of genetic information and the development of the different techniques, *Drosophila* sperm has become one of the most useful materials for quantitative irradiation work in which the real mechanism of irradiation effects is sought. This material allows observation on the lethal effect of the X rays and also on carry-over effects of radiation to succeeding generations. These effects have been shown to be due to the capacity of the X rays to alter the inheritance units, the genes, and to break and rearrange chromosomes.   These effects in *Drosophila* may be studied by both genetic and cytological methods (*26,41*).   Other material as, for instance, corn pollen has equal advantages in that the cell structure is well understood from a cytological viewpoint and each pollen grain is separate and can be followed through as a separate entity.   But

corn has the disadvantage of a longer generation time and a less easily handled cytological technique. Materials like the pollen grains of *Tradescantia* (*48,50*) have proved particularly useful in the study of chromosomal rearrangements due to radiation. These pollen grains have the advantages of being good cytological material; the cytological patterns of cell division, as in mitosis and meiosis, are well understood and standard. The disadvantages appear when the results require progeny studies.

Other materials, like bacteria (*24,25,27*), have the advantages of being single cells, of discreteness, and of high rate of reproduction. They are quite useful for studying the effect of radiation on survival, provided the species is such that the cells do not form chains or clumps. They may be used for the study of problems of frequency of phenotypic mutations, but they have the difficulty that very little is known about their method of chromosome reproduction, with the consequence that interpretation of the results on a comparable basis with those in other forms, in which cytological behavior is known, may lead to false conclusions. Viruses, as the plant viruses of the tobacco mosaic group (*23*), are very useful for X-ray study in that, like bacteria, they may be plated out as individuals and the changes observed in them noted and carried on as genetically different strains. They are useful for the study of the lethal effects of different types of radiant energy and for determining size of entities in the organism and comparing the sizes to those obtained through other methods. The viruses, however, have the same difficulties as bacteria, in that their mode of reproduction is not understood sufficiently to interpret many of the results. In the study of mutations, both natural and due to radiant energy, the bacteriophage types have been useful (*16*). They have given real light on the mechanism of mutation through the changes of successively mutated strains. Again, the difficulty is lack of knowledge of the mechanism involved in the transmission of their inheritance from one individual generation to the next. In the study of mutation frequencies and the types of mutations produced by radiant energy, the different fungus forms have proved useful. This is particularly true of the bread mold *Neurospora* (*7,37*), and of yeast (*38*), in which the cytological arrangements of the reduction division are such as to return all of the products of the maturation divisions for evaluation of the inheritance they receive.

Radiation studies on multicellular organisms have been useful for examination of particular physiological problems, for instance,

the effect of the selective destruction of given types of tissue or tissue
cells on immune reactions. Thus, the mouse of a particular genetic
strain may have its resistance lowered considerably if it is irradiated
with X rays. Such observations throw light on immune phenomena.
On the other hand, the experimental material is defective for the
study of the mechanism by which the X rays produce their effects,
for this mechanism ultimately goes down into single cells, which
would be nearly impossible to isolate and study genetically. The
elucidation of such problems requires the types of studies involved in
such work as that on *Drosophila*. The same consideration holds for
the irradiation of tumor cells or cell tissues. The observations are
interesting from a physiological point of view, but for a critical under-
standing of what the X rays really do, are defective in that effects of
the X rays may be on a minute constituent of the cells, as, for in-
stance, the gene, which is in itself really responsible for the observed
physiological result. Similar considerations are important to the
study of the effects of radiation on either embryological development
or the regeneration of host tissues.

Choice of the material for irradiation is of prime importance. By
preference this material should be well marked by genes so that the
effects of these genes on the embryological unfolding of the develop-
mental pattern could be studied. Experiments of this type have
led to interesting results in *Drosophila*, in which it has been possible
to time the separation of tissues, as those of the eye (*22,46*), from the
rest of the track. X-ray absorption is discrete even to particular
parts of the cell. This discreteness becomes quite evident in tissues
undergoing rapid and slow multiplication, the cells multiplying most
rapidly being those that express the effects of irradiation soonest.
Embryological and regenerative studies utilizing X rays may ma-
terially aid in the solution of problems of immune body production,
phagocytosis, hormone production, and other such vitally important
physiological phenomena.

E. TYPES OF DATA ACCRUING FROM X IRRADIATION
OF BIOLOGICAL MATERIALS

In the past, and, no doubt, in the immediate future, because of
the interest surrounding it, many of the observations on X-ray effects
on biological material have dealt with the survival of the given or-
ganisms under known doses of X rays and specified environmental

conditions. These observations have been made on such biologically different materials as plant and animal viruses, bacteria, *Drosophila* sperm, sunflower and mustard seed, mold spores, yeasts, and protozoa. The graph presented in Figure 7 shows some of these data, together with literature references. The logarithm of the per cent survivors is shown as the ordinate of the graph. The abscissas represent the X-ray doses in roentgens. It will be noted that the survival curves for each species are quite different, both in their form and in
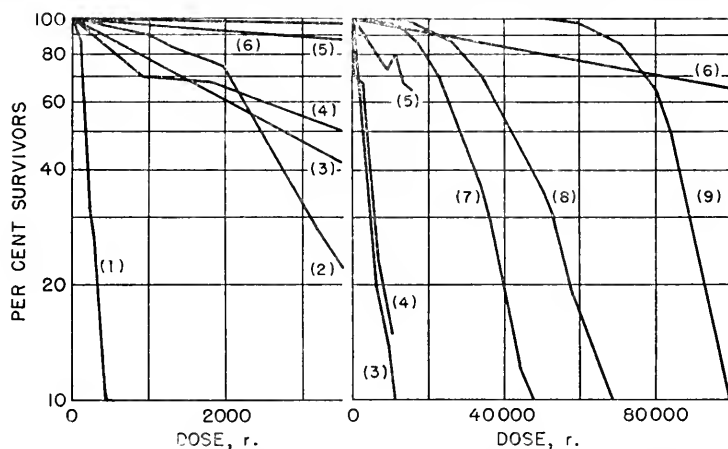
Fig. 7. Survival curves for several species irradiated with known doses of X rays. Forms taken by survival curves illustrate various complexities in X-ray effects requiring Poisson equations of more and more terms to explain them. (1) *Drosophila* eggs (43); (2) yeast cells (53); (3) *Drosophila* sperm (26); (4) colon bacillus (52); (5) *Drosophila* sex-linked lethals (26); (6) ordinary tobacco mosaic virus (23); (7) and (8) *Saccharomyces ellipsoideus* (1.56 and 0.56 A.) (21); (9) *Colpidium colpoda* (10).

the amount of radiation necessary to produce a given lowering of the survival curve. As will be seen later, both these factors are important in the interpretation of the radiation effects and the biological changes they make.

Another type of observation is that of the rate of mutation in the individual genes of an organism when exposed to radiation. Similar data may also be gained for the breakage of chromosomes by X rays, resulting in the formation of inversion, translocation, and deletion in the chromatin structure. Data on lethal mutations in the sex chromosome of *Drosophila* representing gene mutations and chromosome deficiencies follow the simplest type of Poisson equation. Where

the number of genes not affected is equal to y, the original number of genes is N, the X-ray dose in r. units is d, and b is a constant, this equation becomes:

$$y = Ne^{-bd}$$

The exposure of the blood cells of an organism to X rays is known to change the frequency of the leucocytes in the blood. These
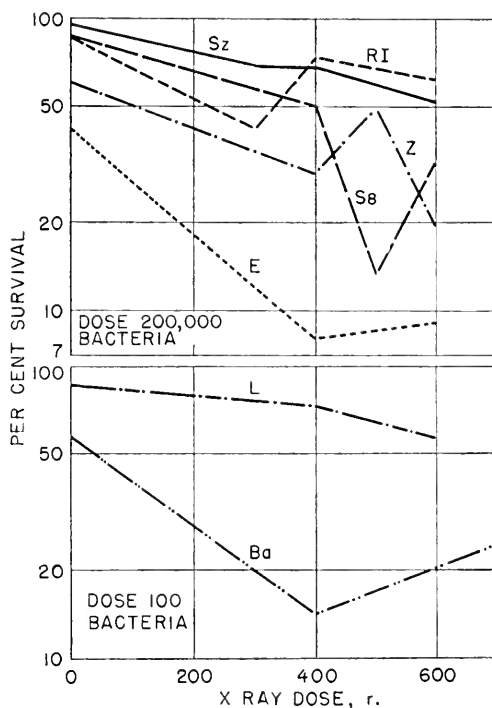
Fig. 8.   Effect of X radiation in lowering the resistance of different mouse strains to given doses of typhoid organisms.

changes have pronounced effects on the resistance of the organisms. Radiant energy has been used as a tool, for example, in altering resistance to mouse typhoid disease.   Mice of resistant and susceptible strains have been irradiated and then inoculated with the disease organisms.   A plot of the results is shown in Figure 8.   The data are quite irregular because of the difficulty of getting sufficient material for test.   A general linear trend downward in resistance is evident with increased irradiation.   This would be expected if the X rays

were randomly destroying the blood-forming cells of the body and thus lowering the number of leucocytes available for opposing the disease organisms.

Much interest has been attached recently to the combination effects of X-ray irradiation with other types of energy, as ultraviolet and infrared (*33–35,50–53*). These combination effects have been studied quantitatively in *Tradescantia* and *Drosophila*. Other forms of energy combined with that of X-ray irradiation lead to results that are pronouncedly different from those when X rays are used alone.

These illustrations of the use of X rays in experimental work depend for their interpretation upon the accurate measurement of irradiation dosages and quantitative estimation of the biological effects. The gathering of the data, their statistical analysis, and their interpretation form the keynote to research advance.

F. METHODS OF MEASURING BIOLOGICAL EFFECTS OF X RAYS

1. Virus Materials

Some plant viruses may be irradiated either dry, in water, or in physiological salt solutions. If they are irradiated dry on cellophane, the layer of material is so thin that the difference in absorption between the surface and the cellophane may be so slight that no correction of the dose for differential absorption need be made. On the other hand, when irradiated in solutions, the absorbing power of the solution must be taken into consideration and, if possible, the depth of the solution made so small that its absorption does not become great. The solution for carrying the treated material should be one in which the radiation will not produce a toxic substance. There are some chemicals that on irradiation materially affect bacteria and viruses. In fact, water itself may release oxygen, which can become a source of error in the experiment.

After irradiation and proper measurement of the dose as it is received by the organism, it is necessary to separate the individual bacteria or virus particles so that a count can be made of the original number and also the number that survive the X-ray treatment. Bacterial species should be chosen that do not form chains or clump and thus completely vitiate the X-ray experiment. Similarly, virus particles must be capable of being distributed separately over the test

material for individual counting. The number of bacteria may be determined by smearing a measured sample of the initial solution on suitable agar media in ordinary bacteriological plates. The separate bacteria then grow into small colonies, and at the end of 24 to 48 hours may be counted. Comparison of the numbers of organisms in the untreated sample with the number in the various exposed samples results in estimates of the effect of the X rays.

Certain plant viruses, as the necrotic type of tobacco mosaic (30,31), may be plated similarly on the leaves of a susceptible plant, e.g., bean seedlings. The solution containing the unirradiated or irradiated virus is spread on the surface of the bean leaf by moistening a small pad of cheese cloth in the virus solution and lightly passing this moistened cloth over the surface of the leaf with one stroke only. More than one stroke is undesirable. At the end of three days to a week, depending upon the kind of test, small necrotic spots will appear on the leaves and may be counted. The number of these spots gives an estimate of the number of virus particles within the original solution. On the face of it the methods seem quite crude, but when a comparison between samples of the same solution is made the results are in good agreement. This gives confidence in the method, although some of its success must seemingly be due to the cancelling out of a variety of errors that could arise in the technique. These results demonstrate the desirability of a careful statistical design to eliminate as many of these errors as possible or to measure their effect and take them into account statistically (see Snedecor, 49).

One of the main errors is the difference in the genetic and physiological susceptibility of different plants of the same species. These differences can be overcome by what is called the half-leaf method, where on one side of the leaf one solution is placed and on the other side of the leaf another solution is inoculated. As the two sides of the leaves are quite similar, both environmentally and genetically, the difference between the two lesion counts on the two halves gives a more accurate measure of the ratio of virus particles in the test solutions. A great variety of these design techniques have been worked out and should be investigated by anyone who intends to use this material. The design chosen should be as simple and as free of possible technical error as is adequate to do the work.

With most animal viruses and many plant viruses no known technique is available by which we can plate out individual virus particles. To obtain results for such viruses it is necessary to depend

upon a dilution technique (see *17,28,29,45*). Work with vaccinia virus is of this type. Vaccinia virus for irradiation is extracted from an infected rabbit testis, generally preserved in glycerin, by grinding the testis in sand and diluting it with a 1/10 physiological salt solution, as, for instance, Locke's. After centrifuging this ground solution a portion of it, say, 0.75 ml. of the supernatant, may be transferred to suitable dishes and exposed to the effects of X rays. Dishes for this exposure have been very satisfactorily made out of ordinary Parawax by simply boring out a flat hollowed depression in the wax and putting the virus solution in this depression. The area irradiated should be sufficient so that the depth of solution is slight and correction for absorbing energy reduced to a minimum. Since X rays are more lethal to bacteria than to viruses, the technique if properly handled will keep the material essentially biologically sterile for other organisms than the vaccinia. The reduction in the virus particles due to irradiation may be determined by titration experiments, in which 0.1 ml. of the irradiated or controlled virus suspension is inoculated intradermally into rabbits in separately marked spots.

The rabbit is first shaved, then the body surface marked off in squares; as many as thirty or forty squares may be possible on a good-sized rabbit. In the center of the square, 0.1 milliliter of the virus suspension is inoculated into the corium. Negative and positive takes will be noted. The virus will not spread but will cause a swelling and reddening of the area for a positive take. Negatives will show little swelling. The different treated solutions and controls should be distributed over the surface of the rabbit, either at random or with some prearranged design capable of easy statistical evaluation, so that there is no place where infection interaction interferes with the interpretation of the data. Positive tests show that virus particles are present. Negative tests show that all the virus particles have been destroyed in the sample inoculated. Only this type of data can be obtained from such material. An estimate of the number of virus particles in the treated and control solutions can be obtained by diluting the solutions and getting the dilution at which approximately 30–50% of the inoculation tests are positive and the rest negative. The dilution factor, the percentage of positives, and the assumption that the particles are distributed as in the Poisson series make it possible to estimate the number of virus particles in the original solutions under test. Comparison of these numbers for differently treated solutions gives the X-ray effect.

2. *Drosophila* Genetics

In determining the effects of irradiation on the germ cells of higher organisms, as, for instance, *Drosophila*, the following points are worthy of consideration. Eggs should be from a strain that lays fertile eggs rapidly, at the rate of something like an egg every ten minutes. The hatchability of the untreated eggs should be 95% or more. Eggs should be laid in a limited period of time which is standard for each test. Packard's standard method for such tests is as follows:

Fifty to one hundred females are allowed to lay their eggs on small pieces of black filter paper wet with the juice of fermented bananas and spread thinly with a little banana substance mixed with yeast. In two hours they will usually lay enough for an experiment. At the end of the two hour period the papers containing the eggs are removed and arranged in a convenient way for irradiation. One hour after removal they are exposed to accurately measured dosages of X rays and then placed in a moist chamber at room temperature. These time intervals should not vary. Three hours after the laying period commences, the eggs are in a most sensitive condition. As they grow older, they become more and more resistant to irradiation, so that three hours later they are approximately half as sensitive. This rapid change in sensitivity makes brief exposures necessary. The eggs should be exposed to X-ray beams having a relatively large number of roentgens per minute.

About eighteen hours after exposure the eggs are examined under a binocular microscope with a 15× magnification in order to count the number hatched before the majority. These early hatched eggs were old when laid, and therefore should not be included with those subsequently hatched, since their sensitivity is very much less than that of the young eggs. The error introduced by not subtracting these early hatched eggs is not large when the dose is light, but it may become quite important with a heavy dose. Two days after exposure the sample is again counted and the proportion of the hatching eggs determined. The corrected percentages are obtained by dividing the actual percentage observed by the per cent normal for the controls, this being the normal fertility of the strain under consideration. This percentage should be 95% or better. If it is not, the material and fly strain should be discarded and another one sought. This is not often necessary as an inbred line of *Drosophila* varies little from year to year in its fertility and hatchability. It is important to keep the fly strains used for these experiments in stock bottles having a large excess of food and without overcrowding, as this arrangement makes for good laying females.

Experiments should be repeated on some good statistical design

so that the different factors affecting variability of the results may be measured and taken into account.

Some experiments require the irradiation of only the female inheritance.   This may be done by irradiating the eggs just after the last oogonial division when they are still retained within the body of the virgin *Drosophila* female.   The mature virgin female has 100 to 150 of these egg cells in her ovary. These eggs are in the diploid condition, as reduction division in the *Drosophila* takes place only after the sperm have entered the eggs.   Immediately after irradiation, the females should be mated to active males.   The results will be biased by effects produced by gene or chromosome rearrangements on the maturation divisions. Comparison of the hatchabilities and other changes induced by the different treatments with those of the controls will give a measure of the X-ray effects.

Treatment of the haploid male gametes is fairly easy in both animals and plants, *i.e.*, in *Drosophila* the sperm may be irradiated as they are contained in the adult males; in plants the pollen may be irradiated after it is shed.   The effects of these irradiations may be observed in a variety of ways. Number of eggs hatching, number of larvae that develop to a given instar, and number of adult flies that hatch are all measures that have been used in the past.   For the more direct effects of the radiant energy, gene mutation, chromosome aberrations, etc. offer good quantitative measures of the changes produced.

With pollen the viability may be determined by the growth of the pollen tube on suitable artificial media or by the number of seeds produced when the pollen is used for fertilization on a standard, genetic type female.   Chromosome and chromatid changes in the pollen tube cells may measure the effect on the chromosomes.   Seed and plant characteristics present opportunities for gene mutation studies.

In much of the earlier work multicellular seedlings were irradiated and the effects of the irradiation determined by the growth of the seedlings in the various treatments.   The composite character of this material made interpretation of the X-ray effects difficult.

One of the most satisfactory methods of determining X-ray effects on vital cell structures is that of detecting so-called sex chromosome recessive lethals in *Drosophila*. This category includes several radiation effects, such as gene mutations, small deficiencies in which a vital gene or a series of genes is destroyed in the chromosome, translocation of a particular kind, etc.   The detection of these chromosome

and gene changes is facilitated by special matings and genetic techniques (26,41). Special tester stocks are desirable for the test. A genetic diagram of four generations of matings is shown in Figure 9.
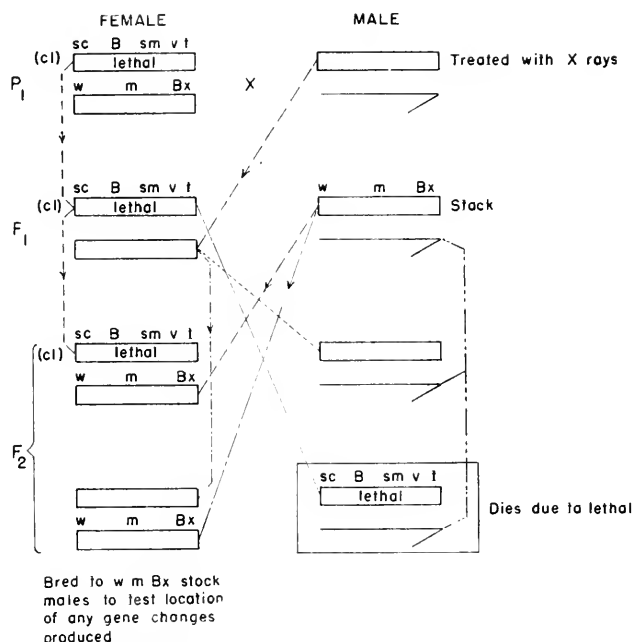
Fig. 9. Diagram of matings for detection of X-chromosome mutations induced by X rays.

In the diagram, the following procedure is illustrated: Sperm are irradiated within the parent male, which is then mated to tester stock female carrying in one X chromosome genes scute(sc), bar(B), small wing(sm), vermillion(v), tan(t), a lethal (not shown)(l), and an inversion to prevent crossing over (not shown)(c); in the other X chromosome, the female carries white (w), miniature(m), beadex(Bx). The females in the progeny carrying the bar chromosome and the X-rayed chromosome are chosen and bred to tester males having w m Bx in their X. If the X rays produce a mutant gene or a lethal in the X chromosome this would be observed in all the males of the next generation (F_2) by the mutant change in phenotype, or, if lethal, by the presence of only female progeny. Location of gene change may be made by mating females showing the beadex (Bx) gene.

A similar method, but applicable to the autosomes, is that for

detecting autosomal sterility factors, gene mutations, lethals, and translocations. Again special *Drosophila* stocks are necessary for making these tests. The general pattern for the test is schematically drawn in Figure 10.

Another method for studying irradiation effects is that by which a loss in the chromosome or a specific gene change at a particular
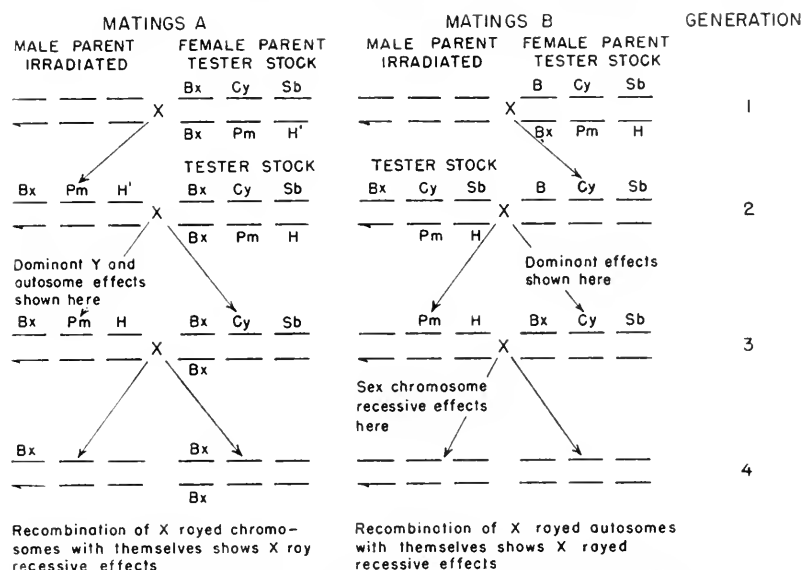
Fig. 10. Genetic diagram for detection of X-ray effect on autosome. For list of symbols, see Bridges and Brehme (8).

locus is detected. Such a test may be carried out by irradiating the dominant allele of a gene and mating it to test organisms that carry the recessive allele in homozygous condition. The appearance of the recessive allele indicates either a deficiency, chromosome rearrangement, or a mutation to another gene in the allelic series. For technique, see Fryer and Gowen (18).

3. Cytological Methods

Cytological methods for the study of X-ray effects depend upon the material exposed to radiation and on the objectives sought. Two general methods are available—one is a paraffin technique, in which tissues for study are imbedded in paraffin, sectioned to a given thick-

ness, placed on slides, stained, and studied microscopically. The process is not simple. Texts on histological technique describe the methods (*19*).

A more rapid method is the smear method (*14*), in which the tissue for the study of the chromosomes is dissected out, placed on a slide in a small drop of acetocarmine or aceto-orcein, and then pressed to a single cell thickness by pressure on a cover glass. Such preparations may be made permanent by replacing the water and acetic acid in the acetocarmine mount by alcohol, xylol, and balsam or euperol.

Since cytological studies have generally sought the chromosome changes produced by the X rays, the tissues studied for these X-ray effects have been confined to those showing good chromosomes. These tissues are the germ cells—in plants, the pollen mother cells, pollen tube nuclei, etc.; in animals, the testis or ovaries (one sex or the other being preferred in a given species), ganglion cells, and salivary gland cells. Detecting such chromosomal changes as chromosome breaks and rearrangements necessitates familiarity with the geometrical figures formed by these chromosomes when they are so rearranged. The extensive nature of these changes makes it necessary to refer to text diagrams (see *3*).

Drosophila chromosomes as found in the salivary glands have the important advantages of large size and consistent arrangement of some 5000 bands along these chromosomes. These bands accurately locate chromosome breaks within the chromosomes. The types of breaks may be identified (*32*). Chromosomes show loss of small sections as deficiencies in the bands. The deletions of large sections are generally lethal, so that these figures are not observed cytologically, although they do contribute materially to the deaths of the irradiated *Drosophila*. Inversions of chromosome sections are observed by the fact that in the inverted sections the chromosomes do not pair properly. Similarly, the interchange of chromosome parts from one chromosome to another is observed as crossed figures where breaks occur. By classifying the data into these categories it is possible to specify the X-ray action as due to one, two, three, etc. absorptions of the X-ray energy to which the material may be exposed.

4. Bacterial Mutations

X-rays are frequently used to speed up the mutation rate of the genes affecting particular characters in the hope of obtaining favor-

able changes.   Two factors are important for success in the study of
this question—the irradiation of fairly large numbers of individuals
and the methods of selecting from a very large population of normal
individuals those few that have the genetic effects sought.   There
are various ways of accumulating mutations in a species.   In *Dro-
sophila* a balanced lethal technique with inversions to prevent crossing
over allows a chromosome to accumulate and hold large numbers of
mutant genes.   After the accumulation process has been carried on
for some time, it is possible to separate the genes by outcrossing the
chromosomes to normal stock and analyzing them for the gene changes
that have taken place.

Another method of considerable importance is that of isolating
desired bacterial mutations produced through radiation effects (*36*).
On culture media bacteria of a given species tend to be selected for
those mutants that are favorable to this saprophytic life.   If the
organism studied is a disease-producing type, this means that the
introduction of these nonpathogenic variants will lower the average
pathogenicity of the bacterial strains.   On the other hand, when the
organism is in its normal host only those bacteria that have virulence
will survive (*54*), and be selected to continue the strain.   These selec-
tion effects can be used in picking out the particularly desirable
types of organisms that may result from irradiation changes in the
genes.

The significance of bacterial mutation to the effect of various
drugs in controlling pathogenicity has been studied in this manner
(*42*).   The bacterial strains are made uniform for a given type by
micropipet techniques in which one organism is picked out, placed on
suitable culture media, and allowed to grow into a colony from which
the strain is started.   The progeny of this single bacterium are
considered genetically alike.   The bacteria may then be treated
with the drug, *i.e.*, any sulfa drug, and shown to be all susceptible.
The object of study may be to establish bacterial strains resistant to
these drugs through mutations produced in them by X rays.   The
original bacterial culture is subdivided into many lines by selecting
individual organisms through micropipet methods or by plating and
isolating single colonies if the species under investigation normally
has only single cells.   A subsample from each of these strains is
plated and examined for mutations (*51*).   If the treatment is adap-
tive, the different strains should show constant, similar variations.
If, however, segregation of the original bacteria has taken place or

if rare mutations are occurring in a random manner, the difference
between the different strains should be highly significant. This
analysis will be useful in determining the prior condition of the culture
under irradiation. Bacteria from the different strains may now be
irradiated with a selected dose. If a given dose of X rays, *e.g.*, 5000
r., is producing the mutants, the bacterial strains should be similar
in the number of mutants observed. Furthermore, if irradiation is
given in increasing doses the increase in number of mutants should be
similar for the different strains for each successive like increase in X-
ray dosage. Experimental methods of this general design allow the
investigator to contrast the hypothesis of so-called adaptive response
of bacteria to the environment with that of randomly occurring muta-
tions.

Radiation experiments may have as their objective the production
of mutated strains that will be more efficient in the formation of end
products. Such experiments were illustrated in the necessity for
better penicillin. Starting from the typical culture of the fungus
producing but few units per milliliter of penicillin under standard
conditions, successive radiations and collection of mutant types vary-
ing in the right direction raised the unit production of penicillin to
761 units per milliliter.

5. Embryological Studies

X-ray techniques have proved useful in many embryological
investigations. Two with somewhat different purposes will be men-
tioned: investigation of the nature of the alterations produced by
radiation within the cell, and plotting the course of embryological
development by means of changes induced by radiation in the develop-
ing organism. Observations on the effects of X rays largely depend
on microscopic study of living cells as radiation progresses. A tiny
beam of radiant energy is directed on the cell. Progressive changes
in the cell occur as irradiation is increased. The first noticeable ef-
fect is an increase in visible fat droplets. This change is not fatal if
radiation is stopped at this stage. The protoplasm becomes turbid
and granular, evidence for denaturation. This stage is, also, not
necessarily fatal if radiation is stopped. Further radiation increases
cell permeability to neutral dyes, *e.g.*, neutral red. Then coagulation
of the proteins actually begins. Protoplasmic streaming is increased,
and then decreased. Viscosity of the cell decreases, and then in-

creases. The cells commence vacuolization, and the mitochondria break up. Chromosomes become fragmented or sticky and clump to form pyknotic nuclei, and cell divisions do not occur. The pH of the cell turns to the acid side, and death, as measured by further cell divisions, occurs.

Observations on X-ray effects on cells have been further refined by using embryological material in the two cell stage and shielding one of the cells from radiation, leaving the other exposed. These techniques have shown that the nucleus and its contained chromosomes are the primary cell structures affected by the X rays. Critical periods in the development have been established. Gastrulation is one in which the organism is sensitive. Similarly, in organ development, the blood-forming system, the nervous system, and the reproductive system are highly susceptible to irradiation effects. On the other hand, tissues that are not in division, as for instance, fat cells, do not show pronounced changes. As development proceeds to the adult, the effects of the radiant energy become less lethal to organ systems with reduced cell division, *i.e.*, the nervous system and the skin.

X-ray techniques have furthered cell lineage studies. Either the sperm or egg may be irradiated to such an extent as to destroy the nucleus but still permit normal fertilization and development to take place. If the sperm receive the nucleus-destroying dose this may lead to parthenogenetic development in which the female inheritance contributes all the future characteristics to the individual. Similarly, by irradiating the eggs before fertilization, fertilizing them, and then allowing development to take place, it is possible to obtain individuals that receive their inheritance from the male parent.

Radiation combined with genetic methods offers further possibilities. Cell lineage studies have been aided by use of an irradiated embryo heterozygous for gene pairs affecting that tissue *(22,46)*. The heterozygous individuals may be irradiated at different stages of the embryological developmental period. The effects sought are those destroying or mutating by X rays the dominant gene to the recessive condition at some particular stage in development. When this occurs, the cells derived from the mutated cell will express only the recessive characteristics. The organ will become a mosaic of two types of cells that may be differentiated by their phenotypes. By this means it has been possible to time in development the stage at which particular anlage are separated from the rest of the developmental

tract, *i.e.*, the eye of *Drosophila* was shown to be laid down as early as twelve hours after development started.   The uses of such methods are yet in their infancy.

X-ray effects on regenerating tissues have quite similar uses. Regeneration studies have one advantage in that if one limb is left intact it acts as a control on the irradiated tissue (*2*, Chapter 12).

G. STATISTICAL TREATMENT OF RADIATION DATA

The statistical methods required in evaluating irradiation data include most of the known methods in the field (*49*).   Three quantitative expressions for the effects of X rays have been adopted. Where the X-ray effect is measured by survival of the organisms, the doses of X rays that cause 50 or 63% of the organisms to die have proved convenient measures.   Where the effects are measured as gene mutations or chromosome rearrangements, a popular means of expressing X-ray action is the percentage of these mutations or chromosome rearrangements per chromosome for a dose of 1000 r. units.   The curves expressing the full range of X-ray effects are of several different shapes depending upon the material irradiated (see Fig. 7).   From an interpretative standpoint one class of these curves is outstanding.   This curve occurs when the relation between X-ray dosage and biological effect follows a Poisson distribution in which a single absorption act is sufficient to produce an observed change (see Fig. 6).   When such data are plotted on the grid, logarithms of the numbers of organisms showing no effect plotted against the increasing X-ray doses, this curve takes the form of a straight line.   Data on inactivation of viruses or bacteria and on gene mutations or certain chromosome rearrangements frequently take this form of curve. The equation representing these results is:

$$n \; = \; n_0 \; e^{-ad}$$

where a is the probability that some specific energy absorption event occur per unit dose in the vital volume, and d is the dose or the number of such energy units to which the biological material is exposed. The survivors are represented by n, the original number by n_0, and e is the natural logarithm.   The mean lethal dose, the dose that will inactivate all but 37% of the material, may be determined directly from the graph.   If a semilogarithmic plot of the X-ray data does not turn out to be a straight line but is a curve showing no changes for a

small X-ray exposure, and finally results in an S-shaped inactivation curve, the single hit or absorption event concept will no longer hold. This means that it takes more than one absorption event in the vital volume of each organism to produce the results. A Poisson equation of more terms will be needed to fit such results, depending on the specific number of absorption events that must occur in the vital volume before the observed effect is produced. Biologically, the number of X-ray absorption events necessary for different materials have been observed to range all the way from one to forty or fifty. The general equation is of the form:

$$ n = n_0 \, e^{-ad} \left[1 + ad + \frac{1}{2} (ad)^2 + \cdots \frac{1}{(r-1)!} (ad)^{r-1} \right] $$

The first term in this equation determines the number of organisms that have not received an absorption event in the vital volume; the second term, those receiving just one; the third term, those receiving just two, and so on to those receiving exactly r absorptions. The biological interpretation of a multiple hit curve is not so satisfactory from a philosophical standpoint as that when only one radiation absorption is sufficient to produce the biological effects.

The Poisson equation has other uses in X-ray research. For certain materials it is impossible to count by customary plating methods the particles affected by the X rays as against those which are not. Dilution methods offer the only easily available means of estimating the number of organisms surviving the different treatments. The radiation experiments yield untreated control and irradiated suspensions of the organisms. The problem is to estimate the number of organisms for each of these treatments. The solution under test is diluted in geometrical series, frequently by multiples of ten, but possibly better for some materials by multiples of five, three, or two. Tests on these diluted samples are run. The place where the most dilute sample shows no lesions represents the case where no self-reproducing virus particles were included in the test solution. The next higher dilution may show positive takes. This represents the case in which one or more virus particles survive in the test solution. By repeated tests the percentage of sterile cultures out of the total number tested is established. The percentage of sterile samples, p, is related to the average number of living organisms as $p = e^{-m}$, where m is the average number of living organisms in the sampling

unit. By plotting these m values a measure of the X-ray effect is obtained and the equation for survival of the biological material against dosage may be calculated. The method is not very precise but is the best available in many instances.

Another method of interpreting X-ray investigations is the analysis of variance (49), where the experiment is analyzed as a whole. Statistical design of the experiment frequently adds greatly to the precision of these interpretations. There is a necessary word of caution, however. X-ray experiments very generally do not meet the requirements of variance analysis, in that the variance of the individual treatment is generally greater for the control and successively less until it reaches its lowest value at the highest radiation treatment. The coefficient of variation rather than the standard deviation tends to be a constant. Such complications make the variance analysis inapplicable. Special methods for transformation of data will be necessary.

H. INTERPRETATION OF RADIATION EFFECTS

In what has gone before, the emphasis has been on the physical effects and viewpoint in radiation experiments rather than on the biological variability of the material. Different schools of thought have interpreted X-ray experiments from either one of these viewpoints. Some have looked on the progressive nature of irradiation effects as a selection of the innate variability of the biological material to the radiation, such that the weakest are killed off early and the strongest survive for the longest times. As the reader has already noticed, the physical viewpoint is that the radiation is distributed more or less at random over the biological material and that it is this random variation that produces the progressive X-ray effects. Evidence is accruing for the importance of both these elements in the interpretation of biological data.

The biological variation of the material has, until recently, been largely postulated rather than proved to exist. The postulate seems good because of the known biological variability of the treated materials as exposed to many other agents. The presumption is that similar variability exists for X rays despite the fact that they penetrate more readily than do the other agents. Recently there has been rather definite evidence indicating that certainly a small amount of biological variability to X-ray radiation does exist. Intraspecies

variability, even though smaller in amount and often negligible in many experiments, would correspond with the interspecies variablitiy already known.

The physical effects of radiation and the hypotheses that may explain the effects have been rather well worked out.   There are some criteria by which the logic may be tested for its applicability to the experiment.   For the one absorption type of exponential X-ray effects the data, when plotted on the arith.-log. grid, will have to be a straight line if the hypothesis is suitable.   The accurate determination of such a line is strong evidence that the effects of irradiation are along physical rather than biological lines.   It hardly seems probable that the biological effect of the irradiations would create such a simple mathematical curve.   However, a further means of proving the physical interpretation exists in the fact that, if the dose is fractionated instead of given continuously, the same biological inactivation occurs.   This should follow on the physical hypothesis but not necessarily on the biological hypothesis of innate variability.   A further test of the effect of radiation is to treat the material with radiations of different types.   Radiation over the X-ray band should behave in a similar manner due to the fact that the larger quanta yield recoil electrons that, in effect, reduce the short wavelength radiation to the long wavelength type before the production of many ions occurs.   This reasoning will not hold for densely ionizing particles, as α rays, which produce distinctly different quantitative results from those of X rays of medium wavelength.   These effects would not be expected on general biological grounds.

The fact that inactivation curves do not behave strictly according to the simple exponential relation is not necessarily evidence against the physical theory that a single absorption act is sufficient to produce the observed biological results.   Several factors are significant in X-ray experiments.   If the bacteria or viruses should be treated and be clumped so that it is impossible to get survival values on single organisms, there will be some divergence from the linear relation expected.   A second factor that may produce results discrepant from those expected is the shielding of biological material from the radiation by other substances that may surround it.   The effect of such shielding could be particularly marked with long wavelength X rays.   Another important factor in causing this curvilinearity is that the X rays may create toxic products in the solution irradiated, and these products in turn affect the particles in solution, rendering

the observed results different from those expected from direct X-ray
absorption.   Careful purification may sometimes avoid these diffi-
culties or at least trace them down to substances in the solution ir-
radiated other than those of the biological material under investigation.

It was pointed out earlier that one of the important effects of
radiation is that of altering the inheritance substances, the genes.
These changes occur at low doses and are impossible to detect micro-
scopically when cells are viewed as they are irradiated.   It is well to
have these facts in mind when interpreting microscopic investigations,
for the microscopic symptoms of irradiation effects may be secondary,
induced by the genes and their changed chemical reactions within
the cells rather than as a direct effect of the X rays.   The effect of
the gathering of fat droplets, the denaturization of the protein, etc.
may be a secondary reaction instead of the sought-for primary effect
of the radiation.

The particulate nature of the irradiation effects seems quite clear
when one examines the genetic chromosome data.   These changes
are either visible or not, as the case may be.   However, the inter-
pretation of just what these changes are and how they are caused is
not very clear.   As mentioned earlier, it has been known for some
time that some disinfectants will give the simple exponential curve
for the inactivation of bacteria with an increasing dose of the dis-
infectant.   This is presumably a chemical process.   So, too, is the
action of radiant energy in the last analysis.   The importance of
these chemical changes to the observed effects is becoming clearer with
the demonstration that certain chemicals may also produce gene
changes and chromosome breaks and have these gene changes and
breaks follow the exponential trend with increasing dosage of the
reacting agent (6).   Mustard gas and dibenzanthracene (12) just
recently have been shown to produce mutations and to cause chromo-
some breaks and deficiencies.   The biochemical specificity of these
changes makes for even greater interest in the relation of X-ray
irradiation to the biological effects and their interpretation.

References

GENERAL

1. Clark, G. L., *Applied X-Rays*, 3rd ed.   McGraw-Hill, New York, 1940.
2. Duggar, B. M., ed., *Biological Effects of Radiation*, 1st ed.   Vols. I and
    II, McGraw-Hill, New York, 1936.

3. Lea, D. E., *Actions of Radiations on Living Cells*, 1st ed. Macmillan, New York, 1947.

4. Compton, A. H., and S. K. Allison, *X-Rays in Theory and Experiment* (2nd ed. of *X-Rays and Electrons* by A. H. Compton). Van Nostrand, New York, 1935.

5. Glasser, O., E. H. Quimby, L. S. Taylor, and J. L. Weatherwax, *Physical Foundations of Radiology*. Hoeber, New York, 1944.

LITERATURE CITATIONS

6. Auerbach, C., J. M. Robson, and J. G. Carr, *Science*, **105**, 243 (1947).

7. Beadle, G. W., in *Currents in Biochemical Research*, D. Green, ed., Interscience, New York, 1946. "Genes and the Chemistry of the Organism," in *Science in Progress*, Fifth Series. Yale Univ. Press, New Haven, 1947.

8. Bridges, C. B., and K. S. Brehme, *Carnegie Inst. Wash. Pub.*, No. 552, 1 (1944).

9. Condon, E. U., and H. M. Terrill, *J. Cancer Research*, **11**, 324 (1927).

10. Crowther, J. A., *Proc. Roy. Soc. London*, *B***100**, 390 (1926).

11. Crowther, J. A., *Brit. J. Radiol.*, **11**, 132 (1938).

12. Demerec, M., *Nature*, **159**, 604 (1947).

13. Demerec, M., *Science*, **105**, 634 (1947).

14. Demerec, M., and B. P. Kaufmann, *Drosophila: Guide for Introductory Studies of Genetics and Cytology*. Carnegie Institution of Washington, Washington, D. C., 1940.

15. Dessauer, F., *Z. Physik*, **20**, 288 (1923).

16. Exner, F. M., and S. E. Luria, *Science*, **94**, 394 (1941).

17. Fisher, R. A., *The Design of Experiments*, 2nd ed. Oliver & Boyd, Edinburgh, 1937.

18. Fryer, H. C., and J. W. Gowen, *Genetics*, **27**, 212 (1942).

19. Gatenby, J. B., and T. S. Painter, in *The Microtomist's Vademecum, A Handbook of Plant and Animal Microscopic Anatomy*. Blakiston, Philadelphia, 1937.

20. Glocker, R., *Z. Physik*, **77**, 653 (1932).

21. Glocker, R., H. Langendorff, and A. Reuss, *Strahlentherapie*, **46**, 517 (1933).

22. Gowen, J. W., *Science*, **70**, 358 (1929).

23. Gowen, J. W., *Proc. Natl. Acad. Sci., U. S.*, **26**, 8 (1940).

24. Gowen, J. W., *Iowa State Coll. Agr. Expt. Sta. Bull. (Annual Reports)*, 153 (1938), 160 (1939), 194 (1940), 172 (1941), 190 (1942), 182 (1943), 208 (1944).

25. Gowen, J. W., *Cold Spring Harbor Symposia Quant. Biol.*, **9**, 187 (1941).

26. Gowen, J. W., and E. H. Gay, *Genetics*, **18**, 1 (1933).

27. Gowen, J. W., and R. E. Lincoln, *J. Bact.*, **44**, 551 (1942).

28. Gowen, J. W., and A. M. Lucas, *Science*, **90,** 621 (1939).

29. Haldane, J. B. S., *J. Hyg.*, **39,** 289 (1939).

30. Holmes, F. O., *Botan. Gaz.*, **86,** 66 (1928).

31. Holmes, F. O., *Contrib. Boyce Thompson Inst.*, **3,** 163 (1931).

32. Kaufmann, B. P., and M. Demerec, *Proc. Natl. Acad. Sci. U. S.*, **23,** 484 (1937).

33. Kaufmann, B. P., A. Hollaender, and H. Gay, *Genetics*, **31,** 349 (1946).

34. Kaufmann, B. P., and A. Hollaender, *Genetics*, **31,** 368 (1946).

35. Kaufmann, B. P., *Genetics*, **31,** 449 (1946).

36. Lincoln, R. E., and J. W. Gowen, *Genetics*, **27,** 441 (1942).

37. Lindegren, C. C., *Iowa State College J. Sci.*, **16,** 271 (1942).

38. Lindegren, C. C., *Bact. Revs.*, **9,** 111 (1945).

39. Luyet, B. J., *Radiology*, **18,** 1019 (1932).

40. Mohler, F. L., and L. S. Taylor, *J. Research Natl. Bur. Standards*, **13,** 677 (1934).

41. Muller, H. J., *Verhandlungen des fünften internationalen Kongresses für Vererbungswissenschaft, Berlin, 1927*, Supplement 1, *Zeitschrift für Induktive Abstammungs und Vererbungslehre,* 234–260 (1928).

42. Oakberg, E. F., and S. E. Luria, *Genetics*, **32,** 249 (1947).

43. Packard, C., *J. Cancer Research*, **10,** 319 (1926).

44. Packard, C., *Radiology*, **25,** 223 (1935).

45. Parker, R. F., *J. Exptl. Med.*, **67,** 725 (1938).

46. Patterson, J. T., *Science*, **68,** 41 (1928).

47. Pinney, L. E., *Iowa State College J. Sci.*, **13,** 269 (1939).

48. Sax, K., and C. P. Swanson, *Am. J. Bot.*, **28,** 52 (1941).

49. Snedecor, G. W., *Statistical Methods*, 4th ed. Iowa State Coll. Press, Ames, Iowa, 1946, p. 485.

50. Swanson, C. P., *Proc. Natl. Acad. Sci. U. S.*, **26,** 366 (1940); *Genetics*, **27,** 491 (1942); *J. Gen. Physiol.*, **26,** 485 (1943); *Am. J. Bot.*, **30,** 422 (1943).

51. Witkin, E. M., *Genetics*, **32,** 221 (1947).

52. Wyckoff, R. W. G., *J. Exptl. Med.*, **52,** 435, 769 (1930).

53. Wyckoff, R. W. G., and B. J. Luyet, *Radiology*, **17,** 1171 (1931).

54. Zelle, M. R., *J. Infectious Diseases,* **71,** 131 (1942).

55. U. S. National Bureau of Standards, Handbook 20, "X-Ray Protection," and Handbook 23, "Radium Protection." Also, *Radiology,* **31,** 481 (1938).

56. Rogers, T. H., *Ind. Radiography,* **4,** 35 (1946).

ELECTRONS, NEUTRONS, AND ALPHA PARTICLES

L. H. GRAY, *Radiotherapeutic Research Unit of the Medical Research Council, London*

A. PLACE OF IONIZING RADIATIONS IN NATURE AND RESEARCH

1. Ionizing Radiations

Electrons that have passed without loss of energy through a medium having some degree of periodicity in its molecular arrangement may, by the diffraction pattern of the emergent beam, yield information concerning the structure of the medium. This very important biophysical application of electron beams is closely allied to the study of X-ray diffraction. More frequently, electrons lose energy when they are deflected from their path, and this energy which is transferred to the medium is very effective in bringing about both chemical and biological changes. In this chapter we shall discuss certain features of the loss of energy by electrons and other ionizing particles that have an important bearing on the interpretation of these changes, and describe techniques and procedures for their study.

α particles and swiftly moving electrons, or β particles, as they are usually called, are capable of ionizing and exciting every kind of atom through which they pass. The particles are referred to as ionizing radiations because they are generally detected and measured by means of the ionization they produce, but it is important at the outset not to overlook the possible biological significance of the excited molecules, which in fact, outnumber the molecules that are ionized.

Neutrons are incapable of interacting either with atoms as a whole or with individual electrons, and are therefore not primary ionizing particles. In the course of their passage through matter of all kinds, however, neutrons give rise to ionizing particles by collision with atomic nuclei that they either set in motion or disintegrate. When matter as rich in hydrogen as the living cell is exposed to a beam of fast neutrons about nine-tenths of the total ionization is produced by hydrogen nuclei or protons, recoiling from the simple impact of a neutron. From the point of view of biophysical research, fast neutrons are therefore in a broad sense analogous to X rays. Each is a nonionizing radiation that gives rise to an ionizing radiation throughout the bulk of the irradiated medium. The energy of the X-ray quantum is converted into electronic energy and that of the neutron predominantly into the energy of recoil protons, which, in the classification of ionizing particles to be discussed presently, occupy a position between electrons and α particles.

Neutrons that have an energy equal to the mean energy of agitation of a molecule at room temperature are called thermal neutrons. Nuclei recoiling from impact with such neutrons have of course insufficient energy to ionize or even to excite other atoms.   Each thermal neutron, however, ultimately becomes incorporated into the structure of a nucleus, which it renders unstable.   The subsequent nuclear readjustment is accompanied by the emission of γ radiation, by nuclear disintegration, or both.   Such interactions are highly specific and therefore their biological effect may depend critically on the composition of the tissue irradiated.   Moreover, special means have to be adopted for the production of beams of thermal neutrons in sufficient intensity for biological research.   The unscreened nuclear fission "pile" is one such source.   Thus, although the passage of thermal neutrons through living matter does not give rise to ionizing particles essentially different in any respect from electrons, α particles and the particles generated by fast neutrons, it will be convenient to consider them separately.

Neutrons that have an energy of a few electron volts are more energetic than thermal neutrons and are often referred to as epithermal neutrons.   From the biophysical standpoint their distinguishing feature is that they generate recoil atoms, which would be expected to give rise to an abnormally high proportion of excitation relative to ionization in the molecules through which they pass.   For this reason, it would be of great interest to investigate the biological effects of such neutrons.   No such investigations have so far been reported, doubtless on account of the technical difficulties that would be encountered in planning such an investigation.

In this chapter the term fast neutron will be reserved for neutrons of energy greater than about 500 e.kv. (electron kilovolts).   It will be understood that what is said of these neutrons will require progressively increasing modification as the energy of the neutrons under consideration falls from this value to that of thermal neutrons through the intermediate regions of slow or epithermal neutrons.

2. Effects of Ionizing Radiations on Living Cells

For the most part electrons, neutrons, and α particles have been used in order to obtain a better understanding of the mode of action of ionizing radiations on living cells rather than applied as tools for the investigation of purely biological problems.   If delivered in sufficient dose these radiations interfere with almost every known function of

the living cell. Change in cell morphology, and particularly changes in chromosome structure, which are brought about by relatively small doses of all types of ionizing radiation in both plant and animal tissues, have been intensively studied. Much less is known of the influence of ionizing radiations on the chemical processes that take place within the living cell. Nevertheless, within limited fields of study there are certain well understood broad principles, very ably expounded by Lea (1), so that a worker taking up this study may reasonably expect to enlarge his knowledge of fundamental processes in normal cells. In certain respects radiation may now be reckoned as an agent capable of producing a controlled disturbance of vital processes, and since its entry is unimpeded by cell membranes it has some advantages over all chemical agents.

In this section it will only be possible to touch upon a few aspects of radiobiology that may be of interest to those not primarily concerned in the investigation of the mode of action of ionizing radiations.

3. Ionizing Radiations as Part of the Environment of Living Organisms

Living organisms of all kinds are continually exposed to ionizing radiations arising from a number of sources. These include trace concentrations of radioactive substances contained within the organism, from the radioactive content of the earth's crust, the ocean, or the atmosphere, or from cosmic radiation that enters the earth's atmosphere from outer space and penetrates to great depths through earth and water.

Radioactive elements are very widely dispersed in minute amounts throughout the whole of nature. They are derived from the parent elements uranium and thorium, and to a lesser extent from protactinium, which are present in measurable amounts in all types of rock. The γ radiation from this radioactive material present in the earth's crust produces about 3 ions per cubic centimeter per second in air over land areas. The γ ray intensity over the sea is negligible by comparison, since the radioactive content of sea water is in general about a thousand times smaller than that of an equal mass of the earth's crust. The radioelements from the rocks very slowly pass into solution in the soil moisture and then into plants. The gaseous radioelements, radon and thoron, diffuse into the soil air, and are carried into the lower atmosphere. Thoron, having a half-life of less than a minute, remains very localized. Radon, which has a half-life of 3.5 days, is carried by winds for immense distances over land and sea. The ionization arising from the

radon and its decay products in the atmosphere at sea level averages nearly
double that due to the radiation from the earth's crust, and is rather more
than the combined effect of γ radiation and cosmic radiation. This ioniza-
tion is mainly produced by α radiation. The radon content of stagnant
underground air or soil air is very much higher (2).

All the types of ionizing particle under consideration in this chap-
ter contribute to the total ionization, but the proportional contribu-
tions will vary greatly from one organism to another depending on its
size and location.   For example, over the land α rays will contribute
most to the ionization of unicellular airborne organisms, or the most
superficial cells, either on the body surface or in the respiratory tract
of larger organisms. Roots may be irradiated about equally by α
particles and electrons.   In general, electrons will contribute most of
the ionization to the deeper layers of large organisms, but in addi-
tion there will be some ionization due to neutrons and to other rays
that are so far only known to us as components of cosmic radiation.
Since the discovery of cosmic rays there has been speculation as
to the possible biological effects of continuous exposure to this
radiation.   In open air at sea level the majority of cosmic ray ioni-
zation is probably produced by electrons of very great energy, but
from time to time there occurs a local release of enormous amounts
of energy in the form of a "shower" of particles of many different
types.   The frequency with which an organism will be exposed to
such particles will be greatly increased if it is placed immediately be-
neath a 2 inch slab of lead, more especially if the experiment is carried
out at high altitudes.   Figge (3) observed that the latent period re-
quired for tumor induction in mice injected with methylcholanthrene
was significantly shorter in mice placed beneath a thick slab of lead
than in controls not kept beneath lead, and concluded from these ob-
servations that cosmic ray showers are cocarcinogenic.   The experi-
ments touch upon a subject of such immense importance that they
need to be repeated and extended.

The total effect at the surface of the earth of all the components
of cosmic radiation, of which the "showers" are only a small part, is to
produce 1 or 2 ion pairs per cubic centimeter per second.   When
contributions from radioactive substances in the air are included the
total is only around 10 ion pairs per cubic centimeter per second.
This is an extremely low rate of ionization.   It corresponds to an X
ray dose rate of 5.10^{-9} r. per second and to the production of only 1
ion pair in a cell 12 μ in diameter once a day.   Cosmic rays alone

would only ionize such a cell about once a week.   Bearing in mind
that some changes in the living cell, including the production of gene
mutations, are almost certainly produced by all types of ionizing radiation, and as regards each individual radiation in strict proportion to the
administered dose, whatever the duration of exposure, it remains a
matter of the greatest interest for future research to establish whether
or not these infrequent ionizations of the cell have played some part
in the course of organic evolution.

4. Ionizing Radiations as a Health Hazard

In recent years, particularly in America, many investigations
have been directed to discovering what changes are produced in animals exposed for long periods of time to very low dose rates of radiation of various kinds, in order to ascertain the hazards to which those
working with ionizing radiations may be exposed.   The lowest dose
rates so far investigated have been about 100 times as great as those
referred to in the last paragraph. γ rays, β rays, neutrons, and α
particles were used in some of these investigations (4).   The various
species of animals examined showed characteristic types of response.
In most cases some significant departure from the normal, e.g., decrease in life span, decrease in testicular weight, increase in body
weight, and hematological change, was observed in animals exposed
throughout life to 2.2 r. per day of γ radiation.   Carcinogenic effects
referred to in the next section (A5) were observed at still lower dose
rates.

It is of great importance to those engaged in research with ionizing radiations to know what is the maximum intensity of radiation
to which they may be exposed without incurring serious risk of injury
to health.   The only recommendations of international standing regarding the maximum permissible daily exposure to personnel refers
to X and γ radiation.   It is that adopted by the International Congress of Radiology in 1928 and is to the following effect: "The evidence at present available suggests that under satisfactory working
conditions a person in normal health can tolerate with impunity exposure to X rays and radium γ rays to an extent of about one international roentgen (r.) per week."

The evidence referred to in the recommendation has been discussed by Henshaw (5), who outlines the confused subsequent history
of the subject arising out of pronouncements by national committees

and by individuals. The latest recommendation of the British X-Ray and Radium Protection Committee (1943) is identical with that given above, which was essentially a British recommendation in origin.

Assuming a five day work week, the maximum permissible exposure is 0.2 r. per day. In the 1936 report of the United States Advisory Committee on X-Ray and Radium Protection and permissible daily exposure is reduced to 0.1 r. per day and this figure has been in force in United States atomic energy projects. In the Canadian atomic energy project the figure has been further reduced to 0.05 r. per day.

The following are the maximum permissible doses of other ionizing radiations received in any one day, which have been adopted at one American atomic energy station as being, to the best of available knowledge, consistent with a daily exposure of 0.1 r. of X or γ radiation.

Alpha particle	0.01 r.e.p.
Beta particle	0.1 r.e.p.
Gamma ray	0.1 r.e.p.
Fast neutron	0.02 r.e.p.
Thermal neutron	0.05 r.e.p.

The r.e.p. (roentgen equivalent physical) may be defined as the dose of radiation that will result in dissipation of 84 ergs per gram of tissue (Sect. D1). These figures are given only as a guide. They have no international authority and their validity would probably be questioned by some of those who have studied the subject most deeply. All that can safely be said is that any revision is likely to be in the direction of a further reduction in the allowed exposures.

5. Ionizing Radiations as Carcinogenic Agents

Recent investigations (4) have provided abundant evidence of the power of ionizing radiations to produce tumors. Single exposures to the skin of rats of about 4000 r. of β radiation from P^{32} were found to induce a multiplicity of skin tumors including almost every known pathological type varying from papillomas to the most malignant kind of carcinoma (6). The incidence was less in mice and not significant in guinea pigs and rabbits. Repeated small daily exposures did not produce tumors in any of the experimental animals but it is well known that in man chronic exposure of the skin to β radiation is carcinogenic. Whole body exposure to both γ rays and neutrons

significantly raised the incidence of lymphoma in mice (7). Whole body exposure to γ radiation also resulted in a raised incidence of lung tumors, ovarian tumors, and tumors of the mammary gland in mice (8). Radiation-induced tumors were also observed in guinea pigs and rabbits, two species in which spontaneous neoplasms are rare. The incidence of ovarian tumors in mice was particularly striking as this was already evident when the dose received each day was only 0.1 r.

It has long been known that bone sarcomas in man, as well as in experimental animals, may result from the fixation in the skeleton of quantities of radium of the order of a few micrograms and it has also been suspected that lung cancer among the workers in the uranium mines at Joachimsthal is an industrial disease resulting from exposure to the radiations from inhaled radioactive material. In neither of these cases is it possible to say with certainty whether the α rays or the β rays are the effective agent.

Lisco, Finkel, and Brues (9), however, have demonstrated that a high incidence of osteogenic sarcoma may result from chronic exposure to the β rays of Sr^{89} and that Pu^{239}, which emits only α rays, frequently gives rise to malignant fibrosarcomas at the site of injection. There can thus be no doubt that both types of radiation are carcinogenic but little is known regarding the relation between the radiation dose and the tumor incidence. It is noteworthy, however, in the experiments in which Sr^{89} was administered, that although there was a latent period during which no tumors were seen, there was no evidence of the existence of a threshold dose rate below which tumors were not produced at all. Moreover, the proportion of animals in which tumors were produced was found to increase as the square of the time from the end of the latent period. If the same is true in man, significant tumor incidence may result from chronic exposure to very low dose rates. For this practical reason, if for no other, the subject is one that requires much more investigation and it is highly desirable that both α and β radiation should be employed for such investigations, and that sources of radiation should be used that deliver well defined doses of radiation to the tissues under investigation.

6. Ionizing Radiations as Mutagenic Agents

Exposure to X rays and other ionizing radiations may cause mutations at a rate greatly exceeding the spontaneous rate. This was

first clearly demonstrated by Muller in 1927 *(10)* in *Drosophila melanogaster* (see Chapter XIV).

This mutagenic action of radiation has been observed in all the organisms investigated, namely, viruses, bacteria, fungi, liverworts, flowering plants, a variety of insects, and mice. X rays have commonly been used as a source of radiation but sex-linked lethal mutations have been produced in *Drosophila* by exposure to neutrons *(11–15)* and to α radiation *(16,17)*. The chance that a given dose will produce a visible mutation is not only of the same order for a variety of transitions between different allelomorphs at the same locus in *Drosophila*, but, as Gowen *(18)* has pointed out, it is not greatly different in organisms as dissimilar as viruses, bacteria, and insects (*Drosophila*), which contrasts with the fact that estimated spontaneous mutation rates in different species are roughly inversely proportional to life span.

It is estimated that in *Drosophila* an exposure of something like 200 million r. is needed in order that there should be produced in half the irradiated genes a mutation whose phenotypical expression is a change in eye color from "white" to "eosin," and about 10 million r. for that corresponding to the change from "wild" type to "forked bristle." These are the extremes listed by Lea *(1)*. The average value for three bacterial mutations affecting colony appearance is 18 million r. The figure for delayed development of bacteriophage resistance in *Escherichia coli* is 300 million r. In *Drosophila* the induction of sterility by the irradiation of the gonads limits the dose to which the germ cells may be exposed by existing techniques to around 3000 r. A similar limitation would obviously apply to any other type of cell that is rendered completely inviable by irradiation. Bacteria have the advantage in this respect not only of being among the more resistant organisms but also in that up to the highest X-ray doses examined the proportion of survivors is exponentially related to the dose. For example, in strain B/r of *E. coli* used by Demerec and Latarjet *(19)*, 1 in 100,000 bacteria survived a dose of 100,000 r. and among the survivors there were 0.02% that showed mutation to bacteriophage resistance. Several hundreds of new strains of bacteria, *Neurospora*, and various other fungi have been produced in this way, manifesting changes in color, enzymic systems, biochemical characteristics, resistance to bacteriophages, drugs, and antibiotics *(20)*. This relatively recent application of radiation has led to a wealth of new information resulting in considerable advances in our

understanding of the genetic control of biochemical processes. X radiation is convenient for producing these mutations, but it is only one among a number of agents including ultraviolet light, the ordinary mustards, the nitrogen mustards, and other chemicals. Although ultraviolet radiation generally produces more mutations than an equally lethal dose of X radiation it is of the two a less quantitative agent on account of the fact that the absorption of ultraviolet energy depends critically on chemical structure. It is usually impossible to estimate at all accurately the amount of ultraviolet energy that must be absorbed per cubic micron of the material of the gene in order to produce a mutation. Ionizing radiations, on the other hand, dissipate an amount of energy per unit mass that is almost independent of chemical structure. Thus, X-ray experiments of Demerec and Latarjet show that 2.5×10^{10} ergs per gram, of energy in the form of ionization, is needed to produce the mutation that leads to bacteriophage resistance in *E. coli*. Since at least ten times this amount of ultraviolet energy is required, and probably much more, even at wavelengths near the peak of mutagenic efficiency, it seems safe to conclude in this case at least that ionization has a much higher mutagenic efficiency than excitation. Since, however, the efficiency of ionizing radiation relative to ultraviolet light for killing the organisms is still greater, a larger proportion of survivors carrying mutations can generally be obtained by the use of ultraviolet light.

We shall endeavor to make clear in subsequent sections how the use of several different ionizing radiations, for example, fast electrons and α particles, in addition to X rays, can throw light on the question of the size and disposition of the material that has to be transformed— questions that cannot be answered if only one ionizing radiation is employed. Lea and Salaman (*21*) by this means have been led to distinguish bacteriophages such as S 13, which behave as macromolecules, from others such as C 36 and *Staphylococcus* K, in which the genetic material appears to occupy only part of the whole virus, and to be distributed, thus resembling, though on a simpler scale, the structure they had previously postulated for vaccinia virus.

B. PHYSICAL CHARACTERISTICS OF DIFFERENT IONIZING RADIATIONS AND THEIR RELATION TO EFFECTS PRODUCED IN LIVING CELLS

1. Available Sources of Ionizing Radiation

Table I (p. 502) lists some seventeen different sources of ionizing radiation, all of which have been employed in radiobiological investi-

gations. The overriding consideration in the design of any experiment that aims at discovering the processes involved in the action of ionizing radiations must be that the biological structures under investigation should be exposed to a known dose of radiation, expressible in terms of ergs of energy absorbed per gram of irradiated tissue. Problems associated with the measurement of dose and the choice of irradiation techniques that enable this condition to be fulfilled will be discussed in subsequent sections. In general there will remain some choice as to the source of radiation to be used. The kind of information that can be derived from an experiment will depend greatly on the exercise of this choice, which governs the manner in which a given total number of ions formed by the radiation are disposed in relation to biological structures of various shapes and sizes. In this section we shall discuss the way in which one type of radiation differs from another in regard to the distribution of the ionization it produces within an irradiated cell.

2. Classification of Ionizing Radiations According to Mean Spacing of Ions along Tracks of Ionizing Particles

Referring again to Table I it will be seen that while there are seventeen types of radiation listed these give rise to only four classes of ionizing particle, namely, electrons, protons, α particles, and nuclear particles. The ionization track of positively charged electrons, or positrons, is indistinguishable from that of negative electrons. Remembering that particles and protons are the positively charged nuclei of He and H atoms, respectively, it is evident that we are really concerned with only two principal classes of particle, the light, negatively charged electrons, and the massive, positively charged nuclei. Very detailed information about the distribution of the ions formed by these radiations in air has been obtained by means of the Wilson cloud chamber. By means of this instrument a supersaturated condition is produced in moist air just prior to the passage of the ionizing particle through the chamber that results in the condensation of water vapor in the form of a single droplet around each ion. The droplets may be photographed to obtain a permanent record of the position occupied by the ions at the instant of formation. When scaled down approximately in the ratio of the densities of tissue and air, such pictures enable us to visualize the distribution of ions formed by different types of particle within living cells. A few examples will be given (Figs. 2–8) but it will be

TABLE I

Ion Density Produced by Different Ionizing Particles

Radiation	Mode of generation	Mean linear ion density (ions/micron of tissue)	Ionizing particle
	Theoretical min. ion density for any particle..	6.3	
Very high energy β and γ radiation....	20–30,000,000 v. betatron, etc...... / Natural and artificial radioelements..........	8.5	
γ radiation....	Radium screened by at least 0.5 mm. platinum as used in radiotherapy....	11	
X radiation....	"Supervoltage" 1000 kv. installation......	15	Electron
	"Deep therapy" 200 kv. installation.....	80	
	X-ray tubes operated at 30–180 kv.......	100	
	"Characteristic" copper K (8 kv.) X rays.....	145	
	Cyclotrons operating at 12,000,000 v......	290	
	"Characteristic" silver L (3 kv.) X rays....	300	
	Cyclotrons operating at 8,000,000 v....	380	
Neutron radiation....	Aluminum K (1.5 kv.)....	460	
	High-voltage ion tubes / 900 kv. deuterium ions bombarding lithium	840	Proton
	400 kv. deuterium ions bombarding deuterium	1,100	
α radiation...	Natural disintegration of radon.....	3,700	α particle
	Natural disintegration of polonium.........	4,500	
	Artificial disintegration of boron or lithium by slow neutrons..........	9,000	
Atomic rays......	Uranium fission.	130,000	Atomic particle

found most informative to study some of the very beautiful original pictures taken by Wilson and others (22).

Characteristically, electron tracks are thin and tortuous, indicating that the ions are widely spaced and that the electron is easily deflected from its path, while proton and α-particle tracks are thick and straight. The difference between the degrees to which the light and heavy particles are deflected from a straight path is an intrinsic property associated with their mass but is not usually important, since in all but the slowest electrons large changes in direction are infrequent on the scale of cellular structure.   The remaining difference between the light and heavy particles—the spacing of the ions along the track—is not an intrinsic property of the particles at all but a function only of their speed and of the total charge of *either* sign that they carry. Physically both primary ionization and excitation produced in air by a proton and electron of the same speed are practically indistinguishable. The reason for this is well understood in terms of the forces of attraction and repulsion that exist between the charged particles and the orbital electrons of the atoms through which they pass, and justifies the treatment of these two particles as entirely equivalent from the radiobiological standpoint. The α particle, which carries two positive charges, produces four times as many ions per micron of track as a proton of the same speed.   Fission fragments that are multiply charged are more densely ionizing still. The average energy lost by fission fragments per ion pair formed in air, however, is practically the same as for protons and α particles (23).   It appears, therefore, that even in the case of these very densely ionizing particles the proportions of ionization and excitation remain unchanged, so that from a biological standpoint we have only to take account of the compression of much more ionization and excitation into a given length of track.

Evidently, these considerations introduce an immense simplification into the analysis of radiobiological problems for, from the standpoint of the physical and chemical processes taking place within the cell, we can forget the multiplicity of radiation sources and fix attention on a single parameter—the mean linear ion density of the ionizing particles—which is a measure of the spacing of the ionization and excitation along the track of the particles that traverse the cell. This parameter varies with the speed of the particle.   It is least for very fast particles and greatest for particles traveling slowly and nearing the end of their range.   The lower curve of Figure 1*a* shows this varia-

tion for protons and electrons. The upper curve, the ordinates of
which it will be seen are at each speed (except at the lowest speeds)
four times as great as that of the lower curve, represents the variation
with speed of the much greater linear ion density of an α particle.
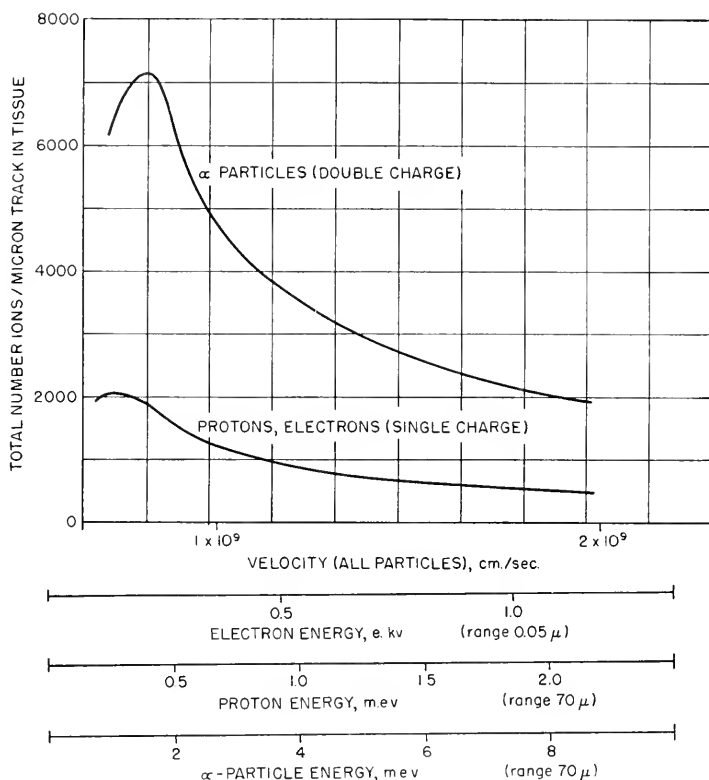
Fig. 1a. Curves illustrating the manner in which the total number of ions
formed by electrons, protons, and α particles per micron of track in tissue varies
with speed and energy of the particle. See also Figure 1b. The curves here
and in Figure 1b were computed from known rates of energy loss by particles
(1, Tables 10–12) for the hypothetical case in which the average energy expended
in formation of a pair of ions in tissue is 32.5 e.v. for electrons and 35 e.v. for
protons and α particles.

At low speeds the α-ray ordinates are less than four times the pro-
ton and electron ordinates because slow α particles capture electrons,
which accompany them during part of their flight and are then lost
again, with the result that the mean charge of the α particle during

this period is less than two units. Four scales of abscissae are given. The uppermost is particle speed and is common to all particles. The remainder are scales of particle energy. From a study of these it is at once evident that the reason we customarily think of protons as densely ionizing particles and electrons as only producing a sparse distribution of ions is that an electron having even a few electron kilovolts of energy is a much faster particle than the most energetic protons and α particles usually encountered.
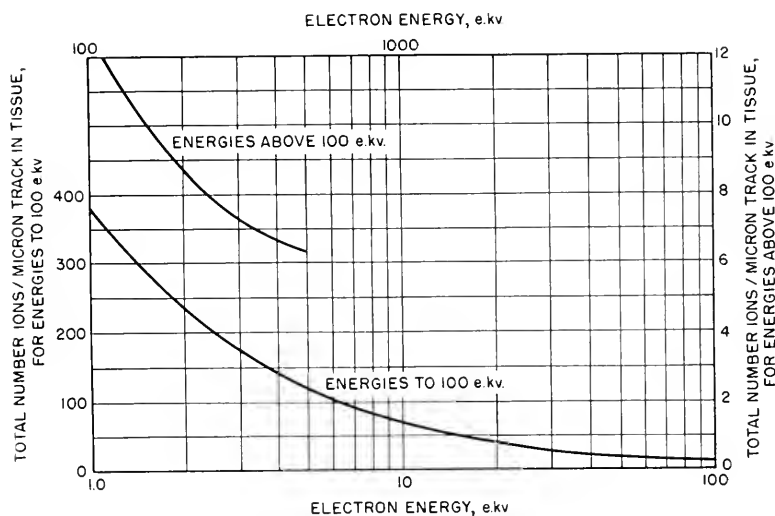
Fig. 1b. Variation of ions formed with electron energy. See Figure 1a for further detail.

Reference to Table I, in which the mean linear ion density of all the ionizing particles generated by each type of radiation is listed, shows, however, that there can be overlapping as between electron and proton ionization, making it possible to test the equivalence of these two particles biologically. Experimental data required for such a test are provided by the observations of Lea, Haines, and Bretscher (24) and of Spear (25) on the lethal action of very soft X rays and neutrons on vegetative forms of *Escherichia coli* and spores of *Bacillus mesentericus*. The dose of α radiation required to inhibit colony formation is about the same for the two organisms but the β-ray dose is nearly 28 times greater for the spores. When relative sensitivities of the two organisms are plotted against ion density for a variety of
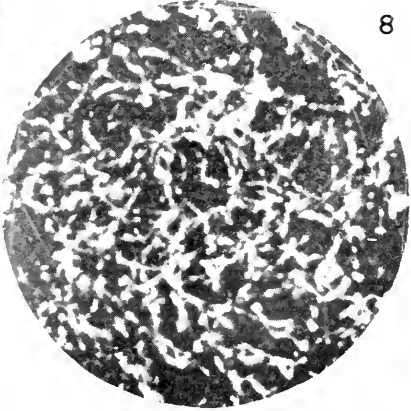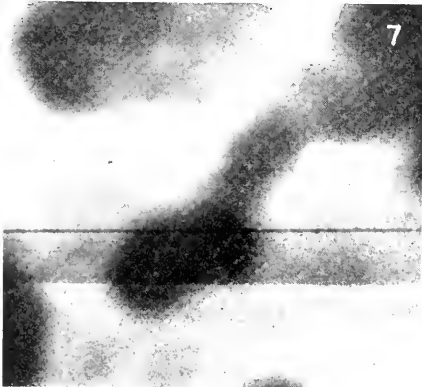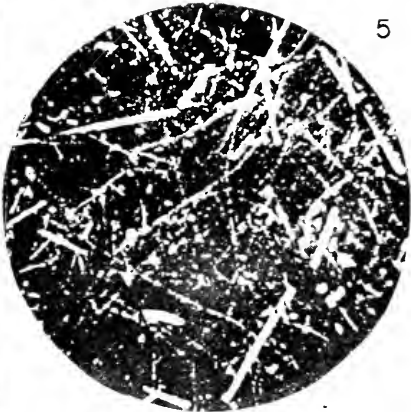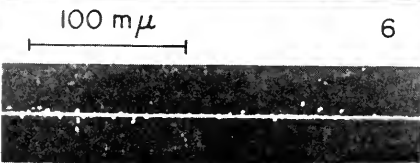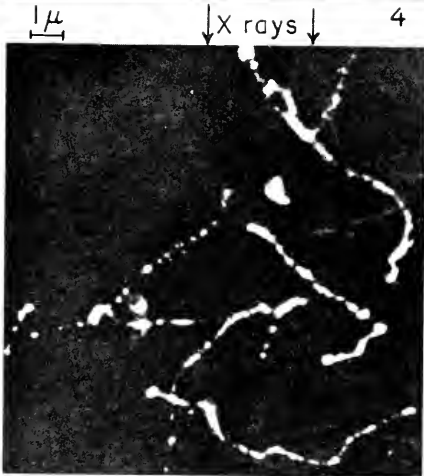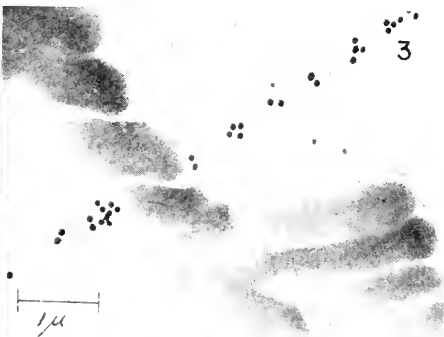
types of X radiation and for neutron radiation the points lie on a
smooth curve, the dose ratio for neutrons closely approximating that
for 1.5 kv. aluminum K X radiation, establishing the equivalence for
this particular biological reaction of proton and electron ionization.

One has always to bear in mind, however, that the full range of a
1.5 kv. electron in tissue is only about 0.1 μ (Fig. 11) and this may
impose a limitation on the effectiveness of the radiation.   For ex-
ample, such an electron could not be expected to break each of two
chromatic threads 0.1 μ in diameter even if these were lying side by
side.   This was verified by Lea and Catcheside (26), who found indeed
not only that exposure of *Tradescantia* pollen tubes to aluminum K
X radiation resulted in a negligible proportion of cells in which both
sister chromatids were broken at the same locus, but also a very low
yield of simple chromatid breaks compared with that resulting from
exposure to higher voltage X radiation.   On the other hand, neutron
radiation giving protons of the same mean linear ion density as 1.5
kv. electrons is very effective in producing both types of chromosome
aberration.   The cloud chamber photographs reproduced in Figures
2–7 illustrate the spacing of the ions along the tracks of different par-
ticles in relation to the size of biological structures.

Figure 2 shows the track of a β particle having an energy of about
1 m.e.v. and traveling with approximately the speed of light (22).
It will be noted that ionization is in the form of clusters separated by
gaps in which no ions are formed.   Each cluster contains an even

Fig. 2.   Wilson cloud chamber photograph of a 1 m.e.v. β particle (22).   Length
of arrow indicates corresponding scale in microns of tissue in this and
subsequent figures.

Fig. 3.   Same β-particle track as in Figure 2 superimposed on an electron micro-
graph of *Escherichia coli* and bacteriophage (22).   Some bacteriophage
may be seen at the right.

Fig. 4.   Wilson cloud chamber photograph of secondary electrons generated by
22 kv. X rays (22).

Fig. 5.   Wilson cloud chamber photograph of recoil protons and electrons re-
sulting from exposure to a mixed beam of neutrons and γ rays from a
cyclotron (28).

Fig. 6.   Wilson cloud chamber photograph of an α-particle track at high magni-
fication (22).

Fig. 7.   Same α-particle track as in Figure 6 superimposed on appropriately en-
larged electron micrograph of tobacco necrosis virus (1).

Fig. 8.   Wilson cloud chamber photograph of electron tracks sufficient in number
to represent the conditions within a nucleus 10 μ in diameter irradiated
by 25 r. of X rays (22).   Same total ionization would be produced by
one α particle passing at random through the nucleus.

number of droplets, as would be expected, since the droplets occur in pairs condensed on the positive and negative ions.   The lateral separation of the ions is out of proportion, having been artificially increased fivefold, by means of an electric field, and will be discussed later. The separation of clusters along the track is about half a micron and we speak of this radiation as producing an average of two ion clusters per micron in tissue.   The frequent occurrence of clusters of more than two ions is noteworthy.   This is true of all ionizing radiations and is due to the fact that the ionizing particle may expel an electron from an atom with all speeds from zero to that with which it is itself moving.   We have to remember also that it may transfer energy to an electron without removing it from its parent atom, *i.e.*, it may merely excite the atom through which it passes.   These events are, in fact, considerably more frequent than the events leading to ionization, so that between each pair of ions we should, to complete the picture, visualize two or three excited atoms, some of which may or may not contribute to the biological effect.   The ratio of the number of atoms excited to the number ionized may well vary considerably from one medium to another, but it is not dependent to any appreciable extent on the nature or speed of the ionizing particle, provided this exceeds that of the orbital electrons of the atom traversed by the particle. Sometimes an electron may be expelled from the parent atom with sufficient energy to produce ions in its turn.   Thus, the clusters of ten ions on the extreme left and also third from the right in the figure each represent one pair of ions produced by the swiftly moving β particle and four pairs produced by the ejected electrons.   On the average a total of three ion pairs are formed for each atom ionized by the primary particle.

These rules cease to be valid when the maximum energy that may be transferred to the orbital electron is comparable with its ionization potential.   In practice, therefore, some deviation may be expected for electrons, protons, and α particles having energies less than about 100 e.v., 200 e.kv. and 1 m.e.v., respectively.   The expected deviation is in the sense of an increase in the proportion of primary to secondary ionization, and an increase in the ratio of excitation to ionization.

In Figure 3 the same β-particle track as in Figure 2 is shown superimposed on an electron micrograph of *Escherichia coli* and bacteriophage taken by Luria and Anderson (*27*).   It will be observed that only rarely are more than two or three clusters formed, within an

object even as large as a bacillus, and, further, that a bacillus will frequently be traversed by a β particle (either primary, or produced by high voltage X or γ radiation) without being ionized at all. It is also evident that an enormous number of β particles, that is, a very high dose, would be needed in order to be sure of leaving a single ion pair in, say, half the bacteriophage particles.

Figure 4 shows tracks produced by 22 kv. electrons resulting from photoelectric absorption of silver K radiation. It will be seen that, although the magnification is five times less than in Figure 3, the ions are much closer together, and finally, as the electron is slowing down, cease to be resolved. This illustrates the general rule that the slower the particle the closer together the clusters and the greater the specific ionization or linear ion density.

Figure 5, reproduced from a paper by Lawrence (28) illustrates the ionization produced by the radiation from a cyclotron. The contrast between the ion density along the recoil proton tracks produced by the neutrons and along the electron tracks produced by γ rays, which accompany the neutrons, is very striking.

Figure 6 shows α-particle tracks at effectively very much higher magnification than in the previous photographs. The bulges on the tracks are some of the larger clusters of ions; a number of δ-ray tracks are visible (see below) (29).

Figure 7 shows one of the α-ray tracks of Figure 6 superimposed upon an appropriately enlarged electron micrograph of tobacco necrosis virus, reproduced by Lea (1). It will be seen that the α particle leaves many clusters inside even such a small particle, by extreme contrast with the faster β particle, which frequently fails to leave a single cluster inside the very much larger *Escherichia coli* (Fig. 3). The difference between X-ray and α-ray ion distribution is also rather strikingly shown in Figure 8. This figure shows actual cloud chamber tracks of 20 kv. electrons sufficient in number to amount roughly to an X-ray dose of 25 r. in a nucleus 10 μ in diameter. The same total number of ions would be produced by a single α particle passing at random through this nucleus.

3. Irregularities in Ion Spacing and Their Biological Significance

The tendency of ions to be formed in clusters was evident in the β-ray photograph (Fig. 2). It is less evident in the other photo-

graphs because the separate droplets are not resolved, but the proportion of clusters of a given size is believed to be much the same at all linear ion densities and has been estimated by Kara-Michailova and Lea (*30*) to be as shown in Table II.

TABLE II
Size Distribution among Ion Clusters (*30*)

Number of ion pairs per cluster	1	2	3	4	>4 (δ rays)
Proportion of clusters containing assigned number of ion pairs....	0.45	0.22	0.11	0.10	0.12

It will be seen that on about 45% of occasions an ionizing particle ejects an electron from an atom that it traverses with insufficient energy to produce further ionization so that only a single ion pair is formed.   Less frequently clusters of two, three, or more ion pairs are formed and on about 12% of occasions the electron is ejected with sufficient energy to produce more than four ion pairs.   Such an electron may itself be regarded as an ionizing particle having a definite track and is customarily referred to as a δ ray.   The distinction between a cluster and a very short δ ray is obviously quite arbitrary.

It is important for the correct interpretation of radiobiological experiments to take these facts into account.   For example, suppose it is desired to calculate the chance that a dose of 100,000 r. of β radiation will ionize an entity having the dimensions of a small plant virus or bacteriophage particle 25 mμ in diameter.   Since the virus is about the same size as the individual ion clusters, we have to calculate the probability that a cluster and not a single ion pair is formed within the virus.   Now, although single ion pairs are the most common form of cluster the weighted mean size of all clusters having the size distribution shown in Table II is three ion pairs.   For all radiations, therefore, the mean number of clusters per unit volume is equal to one-third of the mean number of ions per unit volume. Since 1 r. of β radiation produces 1.7 ions per 10^{-12} gram of virus protein the number of clusters formed is 0.6×10^{12} per gram per roentgen.   Assuming the density of virus protein to be 1.35 gram per cubic centimeter, the average number of clusters formed within the virus particle by 10^5 r. is therefore:

$$10^5 \times 0.6 \times 10^{12} \times \tfrac{4}{3} \pi (12.5)^3 \times 10^{-21} \times 1.35 = 0.66$$

The clusters will be distributed among the particles in accordance

with the Poisson distribution. The chance that a particle of virus protein 25 mμ in diameter will be ionized by 10^5 r. is thus:

$$1 - e^{-0.66} = 0.49$$

The chance that the virus particle is ionized is less than the average number of clusters per particle with the large dose chosen because more than one β particle will ionize the same virus in a proportion of cases.

The chance of ionizing the virus would be almost the same with γ radiation or high voltage X radiation as with β radiation. It would be very different if the virus were exposed to the same dose of α radiation. Table I shows that about 4000 ions per micron are formed along an α-particle track in water. The mean number of clusters per micron is therefore about 1300, so that each α particle will leave some twenty-two clusters within a sphere 25 mμ in diameter. Those virus particles that are ionized at all will be heavily ionized, and correspondingly, for exposure to the same total dose, many more virus particles will be unaffected than was the case with β radiation. α radiation is thus clearly inefficient with respect to any effect that requires only the production of a single ionization within a small spherical volume of virus dimensions. This has been demonstrated by Lea and collaborators (1), as the figures in Table III show. The

TABLE III

Inactivation of Phage S 13 (1)

Radiation	γ rays	X rays (1.5 A.)	α rays (4 m.e.v.)
Inactivation dose[a]................	0.58	0.99	3.5
Target diameter..................	15.5	15.9	16.3 mμ
Unhydrated virus diameter = 16 mμ (hydrated, 18 mμ)			

[a] Unit = 10^6 r.

α-ray inactivation dose is seen to be six times the γ-ray inactivation dose. The figures in the last line show the calculated size of the sphere which is such that there is a chance of 0.63 (i.e., $1 - 1/e$) that such a sphere contain at least one ion cluster. The diameters estimated by the use of each of the three types of radiation agree remarkably well with the estimated size of the phage particle, which makes it possible to assert with some confidence that inactivation results when almost any part of the bacteriophage particle is ionized.

An interesting situation arises when the volume to be ionized is not roughly isometric but is, for example, a spherical shell, a thin disc, a rod, or its equivalent, a large number of smaller spheres. The β- and γ-ray inactivation dose depends only on the total volume within which an ionization is to be produced. The extent of the inefficiency of the α-ray ionization depends on the average path of the particle within the region to be ionized, which is obviously less for the cases cited than it would be for a sphere of equal volume. Such a result was found by Lea and Salaman (31) for the inactivation of vaccinia virus. The γ-ray inactivation dose for this virus was found to be 80,000 r. If the exponential relation between virus inactivation and dose, the observed independence of radiation intensity, and other circumstances are interpreted as indicating that inactivation is brought about whenever a certain region of the virus is ionized, the volume of that region is about 1% of the whole virus and equal to that of a sphere 31 mμ in diameter. If the region were in fact a sphere the α-ray inactivation dose should be eleven times this, or 880,000 r. instead of 211,000 r. as determined experimentally. Lea and Salaman (31) concluded that the sensitive material is nonspherical and probably in the form of some 100 regions each $\sim$ 7 mμ in diameter.

An alternative way of regarding the difference between β and α radiation is to consider that since an α particle produces nearly two ion clusters per millimicron of track it is virtually certain to ionize any structure of dimensions greater than 1 mμ that it actually traverses. If, therefore, one ionization suffices to produce the effect under consideration the α-ray dose measures the cross-sectional area of the "target," by contrast with β and γ radiation, which measure the volume. Lea has treated this problem in very great detail in his papers and in his book (1), where all the necessary physical data are given for the analysis of results obtained with a wide range of ionizing radiations.

It is very interesting to note that the discovery by Preer (98)— that the material "kappa" (which is responsible for the killer properties of certain paramecia) consists of microscopically visible particles—was the direct result of irradiation studies. Finding that kappa was exponentially inactivated, and that the inactivation dose was about 3500 r., Preer concluded that the kappa particles should be at least as large as the largest viruses, and was thereby encouraged to undertake the cytological investigations which lead to their identi-

fication as particles in the size range 0.2–0.5 μ.   Using Lea's analysis
we should conclude that, if the sensitive material is in the form of a
single sphere, it would have a diameter of about 0.15 μ.   If the sensi-
tive material were distributed as a number of smaller spheres through-
out a kappa particle of diameter 0.5 μ, the total value of sensitive
material could be greater than that of a single sphere of diameter
0.15 μ.   The disposition of the sensitive material could be decided
by the use of a second ionizing radiation, probably most conveniently
by the use of neutron radiation.

4. Low Velocity Secondary Electrons, or Delta Rays

Referring again to Table II, it will be seen that on about 12%
of the occasions on which the primary particle ionizes an atom the
ejected electron has sufficient energy to be counted as a separate
ionizing particle (δ ray).   Although for all types of ionizing particle
these δ rays account for only about 20% of the total ionization, there
are a variety of circumstances in which they may play an important
role in radiobiology, as the following examples will show.

First, the δ rays constitute spurs along an α-particle track that
effectively treble the length of the track with respect to any biological
effect that requires only the production of a single ionization.   Thus,
the agreement between the size of the phage S 13 as estimated by
γ rays, X rays, and α rays, is only obtained when the δ rays are al-
lowed for.   In the case of the protons generated by fast neutrons,
e.g. (D–D) neutrons, the δ rays only add about 25% to the effective
length of the tracks.   In the case of β radiation there is no appreciable
increase in effective track length but here other considerations arise.
There are many biological effects, including the breaking of chromo-
somes in *Tradescantia, Vicia*, and many other species of plant and
animal, that are brought about with greater efficiency by neutrons
and α radiation than by β or γ radiation.   From the trend of the
variation in efficiency in passing from α radiation to neutrons and X
rays it would have been anticipated that fast β rays, which will
rarely produce more than one ion cluster in crossing a chromatid
0.1 μ in diameter (cf. Fig. 2), would practically never break the chro-
matid thread.   Kotval (32), however, found no great difference
between the efficiency of γ and X rays, and there is little doubt that
this was correctly ascribed by Lea and Catcheside (26) to the fact
that the δ rays, ejected by the fast electrons, maintain a constant

proportion of ionization in a more effective form. In this instance the δ rays are more effective by virtue of being more densely ionizing than the primary radiation. Conversely, in radiochemical reactions the δ rays along an α-particle track may be more effective by being less densely ionizing than the primary radiation. An interesting example of this is provided by the recent work of Dale, Meredith and Gray (33). The ionic yield for the inactivation of the enzyme carboxypeptidase is ten to twenty times smaller for α than for X rays, a result that may be ascribed to the fact that the highly concentrated radicals formed by the ions along the α-particle track are inaccessible to the enzyme, particularly at low enzyme concentrations. However, contrary to expectation, the α-ray yield did not tend to zero at low enzyme concentrations but to a constant value of about 5% of the X-ray value. This result becomes intelligible when account is taken of the fact that 5% of the total α-ray ionization is produced not along the track of the α particles but along δ-ray tracks, which are not sensibly different in ion density from the secondary electron tracks resulting from exposure to X rays.

For a fuller discussion of the significance of δ-ray ionization in the interpretation of radiobiological experiments Lea's book should be consulted (1).

Lea has computed the energy distribution along δ rays generated per micron of path of the primary particle for electrons, protons, and α particles, using the theoretical formulas of Bethe (34), which describe the loss of energy by charged particles. The experiments of Hornbeck and Howell (35) and those of Shearin and Pardue (36) indicate that the theory is correct for the production of fast δ rays by fast primary electrons. The computed emission of δ rays by fast α particles agrees fairly well with the experiments of Alper (37) but in the case of slow α particles the agreement is poor, the computed emission being greater than that observed. In view of the obvious radiobiological importance of δ-ray ionization it is unfortunate that there should be so little detailed experimental evidence available.

5. Relative Positions Occupied by Positive and Negative Ions at Moment of Formation and Their Subsequent Movement

Ionizing particles very rarely indeed eject an electron from an atom through which they do not actually pass. The primary positive ions, therefore, lie strictly along the geometrical path of the particle. The

particular atoms ionized are picked out at random but the numbers of each kind are roughly proportional to the proportion by weight of this element in the medium.

The figures for the actual number of atoms that would be ionized by a 1 m.e.v. proton crossing a chromatid thread, regarded as a filament of hydrated nucleoprotein 0.1 μ in diameter, are: H, 14; C, 49; N, 16; O, 25; and P, 1, respectively. A sulfur atom would be ionized on an average once in every two transits. The location of the primary positive ions may, therefore, be specified rather precisely. We are much less well informed about the negative ions.

Wilson chamber photographs by Klemperer (38) show that in moist hydrogen gas the negative ions, formed by the attachment of the ejected electrons after they have been brought to rest, are situated at a distance equivalent to about 15 mμ of tissue from the positive ions. Jaffé, from electrical measurements in compressed air and in liquid hexane, arrived at comparable figures for the mean distance separating positive and negative ions in each of these media. No data are available for water but it seems likely that it is at least on the order of magnitude of 15 mμ.

It is important not to overlook the fact that the picture presented of the location of the ions formed by the various types of particle refers only to the state of affairs at the instant of ion formation. Many facts, such as, for example, the reversal by a second irradiation of a gene mutation, itself produced by irradiation, suggest that either the potential energy acquired by a molecule when it is ionized, or the ion itself, can travel considerable distances within a macromolecule. The movement of such energy would certainly repay careful experimental investigation.

When the molecules ionized are small, such as the water molecule, and are formed free in an aqueous medium, their subsequent movements up to the moment of neutralization will be the result of electrical attractions and repulsions, and of diffusion. Since, from the point of view of the interpretation of the biological effects of radiation, we are mainly interested in aqueous solutions, it may well be found necessary to take account, also, of the spontaneous ionization of the medium and the electrical condition at the surface of the larger solute molecules, whose inactivation is of special interest.

The inactivation of pure solutes in aqueous solution has revealed a number of striking differences between the radiochemical yield obtained with β, γ, or X radiation on the one hand and protons or α

radiation on the other, which undoubtedly reflect a dependence of the yield on the linear ion density of the ionizing radiation.   The twenty-fold smaller yield in the inactivation of the enzyme carboxypeptidase when irradiated in dilute aqueous solution by α radiation as compared with X radiation has already been mentioned.   The ionic yield for the destruction of tyrosine observed by Nurnberger (39) using α radiation was nearly thirty times smaller than that observed by Stenstrom and Lohman (40), who used X rays.   Again, very early in the history of the subject Duane and Scheuer (41) made most careful measurements of the decomposition of gas-free water by α radiation and observed a large initial production of hydrogen peroxide and hydrogen, which gave way at larger dose levels to decomposition into oxygen and hydrogen, whereas it now seems clear from recent French work, reviewed by Frilley (42), confirming the much earlier observations of Fricke (43), that the yield of hydrogen peroxide obtained when air-free water is irradiated by X rays is very low.

Evidence is steadily accumulating in support of the view advanced by Weiss (44) that the active agents in most radiation-induced reactions in dilute aqueous solutions are radicals of hydroxyl and atoms of hydrogen derived, respectively, from the positive and negative water ions formed directly by radiation along the tracks of the ionizing particles.   Thus:

$$H_2O^+ \longrightarrow OH + H^+ \qquad H_2O^- \longrightarrow H + OH^-$$

We may presume that this transformation will take place rapidly compared with processes of diffusion so that the initial distribution of OH and H radicals will be sensibly that of positive and negative ions, respectively, at the moment of formation.   Thus radicals of both types will be initially found within a cylinder of radius about 15 mμ around the track of the ionizing particle.   The subsequent movement of radicals may be calculated by the application of the equations of diffusion and of chemical kinetics, as has been done for certain simple cases by Lea (1).   In this connection it is important to distinguish two ways in which the linear ion density of the radiation may influence the chemical changes brought about.   In the first place, since the radius of the cylindrical column of radicals is independent of linear ion density the *average* initial concentration of radicals will increase with increasing linear ion density, so that interaction with solute that takes place in competition with the recombination of the radicals to form water would, on this account, be expected to de-

crease with increasing ion density as observed in the case of tyrosine and carboxypeptidase.

It is, however, only legitimate to consider both types of radical randomly distributed throughout the column as long as the distance between successive OH radicals (always formed near the axis of the column) is comparable with the column radius. Figure 9 shows that while this is true for fast electrons it is of doubtful validity for a 1 kv. electron and is certainly not the case for α radiation. Lea has
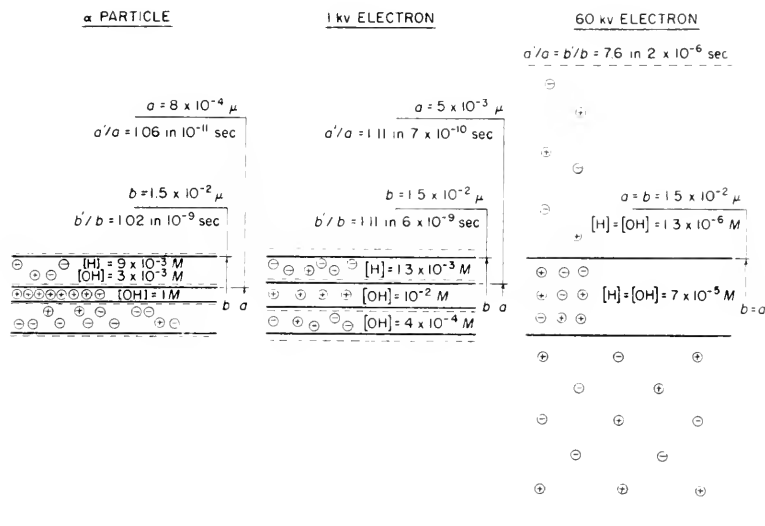
Fig. 9. Diagram showing spatial distribution of hydroxyl radicals and hydrogen atoms, derived, respectively, from positive and negative ions, at the moment of formation by a 60 kv. electron, a 1 kv. electron, and an α particle. Broken lines show expansion of columns during time that 50% of the radicals make one collision. a and a' are the radii of the positive ion column before and after diffusion; b and b' are radii of negative ion column before and after diffusion.

computed the manner in which the concentration of each type of radical varies with distance from the axis of an α-ray track at different times after the moment of formation (see Fig. 10). Noting that concentrations are plotted on a logarithmic scale it will be seen that for a considerable period of time the hydroxyl concentration in the inner column greatly exceeds that of the hydrogen atoms. During this time OH radicals will have made a large number of mutual collisions.

Dale, Meredith, and Gray (*33*) suggest that herein lies the explanation of the high yield of hydrogen peroxide production and of decomposition, observed when air-free water is irradiated with α rays but not observed with X rays and other less densely ionizing radiations.

On the other hand, Allen (*103*) has concluded from studies of the influence of hydrogen gas pressure on the equilibrium concentration of H_2O_2 that results from the exposure of water to intense beams of electrons and deuterons that, at any rate under the conditions of his
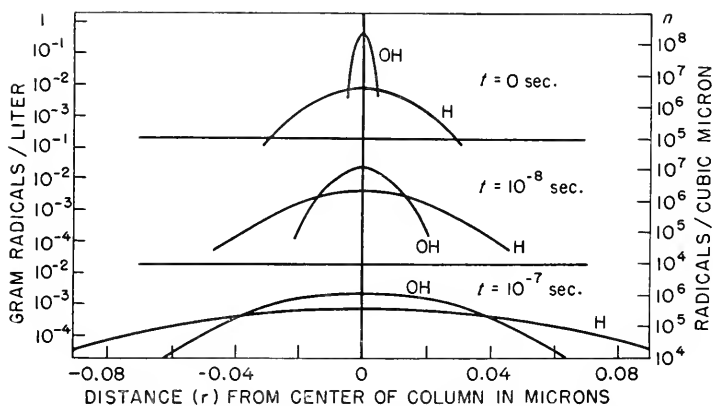
Fig. 10.   Curves showing progressive diffusion of hydroxyl radicals derived from positive ions and hydrogen atoms derived from negative ions formed by an α particle (*104*).

experiments, the poor yield obtained with electrons is mainly due to a back reaction involving molecular hydrogen, which takes place throughout the bulk of the solution and not specifically along the tracks.   Since important constants in Lea's calculations are very imperfectly known at present, the relevance of initial combination of radicals to the interpretation of particular radiochemical reactions cannot yet be properly assessed.   Nevertheless, the differences depicted in Figures 9 and 10 represent a second and distinct aspect of ion density dependence relevant to the interpretation of radiochemical studies.   Furthermore, evidence is accumulating in favor of the view that, even in cases, as for example chromosome fragmentation, in which the visible damage to a particular structure has correctly been ascribed to the action of a single ionizing particle, the damage may

be mediated through radiochemical change either within the structure or in the immediately surrounding fluid. In such cases differences in radiochemical efficiency as between one radiation and another must contribute to the over-all efficiency of the radiation in causing the damage under consideration; indeed, the similarity between the curves relating biological efficiency to ion density for chromosome damage, and some other effects, in widely different organisms, may well result from the fact that these particular biological effects result from the same, or closely related radiochemical changes, brought about with characteristic efficiency by each type of ionizing radiation.

C. SOURCES OF IONIZING RADIATION FOR RESEARCH

From what has been said in the previous section, it follows that the quantitative interpretation of any particular radiobiological exposure will be simplified if the ionizing particles to which the cells are exposed all have the same speed. On account of the limited range of α particles and protons, and of electrons of less than about 100 e.kv. of energy, this is in general possible only if the biological specimen is a unicellular organism that may be exposed in a thin film or on the surface of a culture medium. Cells irradiated within a mass of tissue, whether by α or β radiation from radioactive material present in the tissue or by an external source of X rays, γ rays, or neutrons, are inevitably exposed to inhomogeneous ionizing particles. In this section we shall present data as to the velocity distribution among the ionizing particles from which the mean linear ion density of the radiation may be estimated, as well as of the intensity and penetrating power of the radiation from available sources.

1. Electrons

Electrons of Uniform Energy. The only natural sources that emit electrons of constant energy are a few radioactive isotopes such as Te^{129} that undergo a nuclear isomeric transformation. The charge and mass of the nucleus are unchanged and no particle is emitted from the nucleus but only energy in the form of a monochromatic γ ray, which in a proportion of disintegrations may undergo "internal conversion," *i.e.*, the whole energy of the quantum is transferred to an orbital electron of the disintegrating atom, which is therefore emitted with an energy that is constant and equal to the difference between the energy of the quantum and the orbital energy of the electron. It

has been suggested (45) that the distribution of such elements in tissue might be investigated by an obvious adaptation of standard electron microscope technique.   Whether or not this use of nuclear isomeric transformation proves of value in biological tracer research, it appears extremely doubtful if sources could ever be prepared in sufficient intensity free from isotopes that emit β rays of continuous energy distribution to be useful in radiobiological research.

In principle, the artificial production of electron beams of constant energy requires only a source of electrons, which may conveniently be a heated tungsten filament, a source of high potential, and
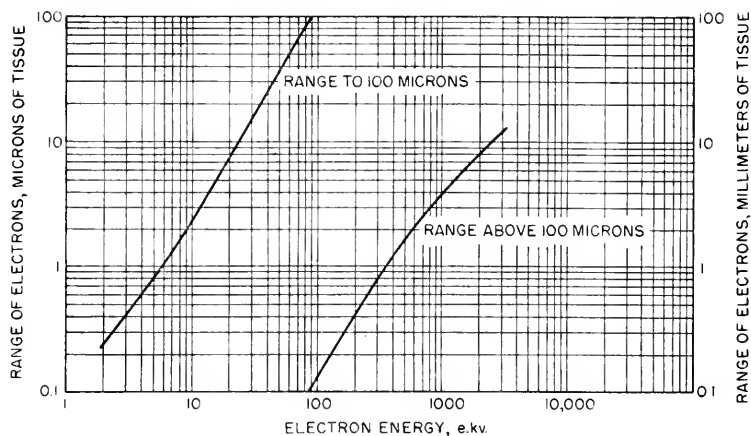
Fig. 11.   Range in tissue of electrons of energy between 2 kv. and 1 m.e.v. Above 1 m.e.v., the range, R, in millimeters, is related to the energy, E, in m.e.v. by $R = 4.8E - 0.8$ (cf. Feather's formula $R = 0.52E - 0.09$, when R is expressed in grams per square centimeter).   For the purpose of estimating rate of loss of energy by electrons the total length of track may be taken to be about 1.4 times the range.

an evacuated tube in which the electrons may be accelerated.   Apparatus of this kind especially designed to meet certain radiobiological requirements in the energy ranges 1–15 kv. and 10–100 kv. has been described (46,47).   After acceleration the electrons pass through a defining slit onto the specimens contained in an evacuated box. Irradiation *in vacuo* is practically unavoidable at the low range of energies since the penetrating power of even a 10 kv. electron is little over a millimeter of air at atmospheric pressure (cf. Fig. 11).   It is possible, however, to make extremely thin windows that will with-

stand atmospheric pressure without imploding, if mounted on a grid
so that the area of unsupported film is very small (48). The two de-
signs of apparatus referred to deliver electron current densities of
0.2–0.5 μa. per square centimeter to exit slits 0.35 × 6.3 mm.[2] The
resultant maximum dose rates to the superficial micron of exposed
tissue varies with kilovoltage from about 5×10^6 r. per second at 1
kv. to 5×10^5 r. per second at 15 kv. with the smaller apparatus,
from 10^6 r. per second at 10 kv. to 2×10^5 r. per second at 100 kv. with
the larger apparatus.   The dose rate is related to the current per unit
area by the following general formula, which is applicable to both
electrons and protons:

$$\text{dose rate} = 3.88 \times 10^7 (i/A) \Delta \text{ r.e.p./sec.}$$

$$= 3.49 \times 10^7 (i/A) \Delta \text{ energy units/sec.}$$

where (i/A) is the current density in milliamperes per square centi-
meter and Δ is the ion density of the radiation in ions per micron of
tissue.

No essential difference in principle is involved in accelerating
electrons to voltages between 100 kv. and 2,000,000 volts.   A number
of forms of high tension supply are commercially available up to 200
kv.   Apparatus to supply higher voltages at constant potentials may
be bought but is to be regarded as a special requirement (49, p. 1391).
The electrostatic generator is a convenient and flexible apparatus for
generating a variety of radiations, including constant potential elec-
tron beams at 1–2 m.e.v.   Currents of 300 μa. may be obtained with-
out great difficulty at 2 m.e.v. and focused to a window 5 cm.[2] in
area, giving a current density of 60 μa. per square centimeter and an
average dose rate of over 10,000,000 r. per second throughout a depth
of about 9 mm. of tissue.   The extent to which the dose rate varies
throughout this depth of tissue has been measured by Trump, Van
de Graaff, and Cloud (50) and is shown in Figure 12.

Apparatus generating electron beams at high voltage should on
no account be approached except under conditions in which every
part of the body surface is protected by sufficient absorbing material
to stop the β rays completely.   It is not always appreciated how
easily β rays are scattered even by air; and in at least one instance
several observers were seriously burned by looking for less than half
a minute and at a distance of several feet at the electron beam as it
emerged into air from a 1.2 m.e.v. generator (51).

Biological specimens should never be placed in close proximity to the window of the tube in which the electrons are accelerated. Not only is the window apt to become warm on account of the electron beam energy it absorbs, but it also is a source of slow secondary electrons. These electrons are the δ rays discussed in the preceding section, but characteristic of the passage of an electron through the material of the window, and since all δ rays commence to travel in the forward direction at the moment of ejection, a cell placed in immediate contact with the window would be exposed to more than half this total δ-ray emission. The complementary proportion of δ rays will, of course, come from the material that supports the cell. Although
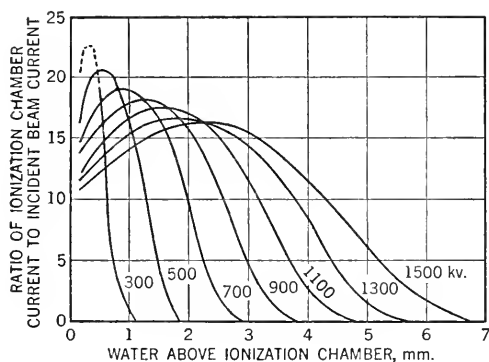
Fig. 12. Distribution of ionization with depth of penetration into water for cathode rays incident with different energies (50).

δ-ray ionization is only about 20% of the total it may completely obscure any dependence of biological effect on the spacing of the clusters along the track of the ionizing particles in any cases in which the high ion density associated with the slow electrons has a high biological efficiency (cf. the production of chromosome structural change discussed in Sect. B2). In certain circumstances it is therefore of interest to reduce δ-ray ionization to a minimum, which may be done by supporting the biological specimen on the thinnest possible film at a distance from the tube window and from all other solid objects. When it is desired to study the effect of the primary particle together with the δ rays that would accompany it as it traverses a mass of tissue, the specimen should be mounted at a distance from the window but with tissue-like material on both sides.

Pulsed Electronic Radiation. Very small currents of 20 m.e.v. electrons have been extracted from the betatron by Kerst and Skaggs (52). The microwave linear accelerator may ultimately prove a

more convenient source of very high voltage electrons. Both sources deliver pulses of electronic radiation.

It is always of interest to extend the range over which a particular variable in any investigation may be studied and, no doubt motivated by this common scientific curiosity, a number of experiments have been carried out with radiation delivered in impulses of extremely short duration. Thus Kingdon, Zahl, Haskins, and Tanis (53) exposed *Drosophila* eggs, *Aspergillus* spores, and wheat seedlings to impulses of 0.3 to 3.5 r. of X radiation delivered in five microseconds, and repeated once a second, but noted no significantly different response from that to the same total dose of X radiation delivered at ordinary dose rates. To produce these X rays an "impulse generator" of the kind used by electrical engineers for testing the strength of dielectrics exposed to electrical surges, and consisting of a bank of condensers, applies the desired voltage to a special design of tube in which instantaneous electron currents of the order of 400 amperes are obtained by autoelectronic emission. If, instead of allowing the electrons to fall on a target, they pass out of the tube through a thin window, much higher radiation doses per impulse can be obtained. Thus, Brasch and Huber (54) report having obtained a variety of unusual biological effects when using electron beams derived from a nominal 3.6 m.e.v. generator giving impulses of 1 microsecond duration and mean current during the impulse of 40,000 amp. The dose of radiation delivered per impulse was not stated but in some cases the specimens may well have been exposed to as much as 100,000,000 r. in one microsecond. Unless steps are taken to expose specimens only to electrons generated during a fraction of the total duration of the discharge, very inhomogeneous electron beams are obtained in this way since the condensers are completely discharged at each impulse.

Since doses of 100,000 r. and upward are generally convenient for radiochemical studies, impulses of high voltage electronic radiation may prove of considerable value in the field of chemical kinetics. Clearly in those cases in which a biological effect stems from a particular ionization or excitation and is independent of changes in the surrounding molecules, impulse radiation will show no characteristic effects. It may well be, however, that some biological changes that are correctly interpreted as derived from the effects of a single ion cluster are nevertheless brought about with greater probability by large clusters than by small. Effects such as chromatid break produc-

tion, which vary linearly with dose and are independent of intensity
for high ion density radiations, and for which biological effectiveness
increases with ion density, may be suspected *a priori* of belonging to
this class.   In such cases impulse radiation may be expected to show
enhanced efficiency.   There is doubtless a limitation both in space
and in time for the cooperative action of ion clusters.   The spatial
limitation will define a minimum dose per impulse, the temporal a
maximum duration of the impulse.

Inhomogeneous Electron Beams.   Almost all our knowledge
of the biological effects of electron ionization has been gained by
exposing specimens to electrons having a continuous distribution of
energies from zero to a maximum value.   This is, of course, true of
tissues exposed to the secondary electrons generated by any kind of
X radiation (see Chapter XIV) and all tissues exposed to the β rays
from radioactive sources.   The great variety of artificial radioactive
sources now available makes it possible to choose continuous β-ray
distributions having almost any desired *mean* energy.   In almost
every case so far investigated the mean energy, E_β, with which the
electron is emitted from the disintegrating atom lies between 30 and
45% of the upper limit E_0 of the β-ray spectrum, which is the quan-
tity generally listed in tables of radioactive isotopes.

The values of E_β for 23 isotopes have been calculated by Marinelli,
Brinckerhoff, and Hine (55) and lie between 120 kv. (Fe^{59}) and 1.39
m.e.v. (Cl^{38}). S^{35} and C^{14} for which E_β would be about 50 kv. and
tritium (H^3) for which it is probably of the order of 4 kv. are not
included in the list.   The value of E_β is the mean value of the energy
of the electrons with which a thin biological specimen would be irr-
radiated when exposed to a source in which both self-absorption and
back-scatter from the source support were negligible.   The corre-
sponding mean linear ion density may be read from Figure 1.   When
a thick layer of tissue is irradiated by an external β-ray source, or
when the radioactive material is incorporated in the tissues the mean
energy of the electrons by which each volume of tissue is ionized is, of
course, a good deal less than the value of E_β for the source in question
because each electron contributes ionization at all energies below
that with which it is emitted.   In this case the mean in which we are
interested can only be properly evaluated when the form of the β-ray
spectrum is known.   A *very* rough value of the mean ion density may,
however, be found by dividing E_β expressed in volts by the product

of 32.5, the energy required to produce a pair of ions, and the full length of track, in microns, of a particle of energy E_β.

Radon contained in a glass or metal bulb of just sufficient thickness to stop all the RaC' α rays provides a very convenient source of β rays, which come from the RaB and RaC, which attains equilibrium with the radon about four hours after the bulb is filled. When specimens about 1 mm. in thickness are irradiated, the γ rays from RaB and RaC may contribute a few per cent of the total dose.

Lea has given useful particulars about the energy distribution among the β rays from such a source. The dose rate at 1 cm. from a point source due to the β rays emitted by RaB and RaC in equilibrium with 1 millicurie of radon may be taken as about 20 r. per minute. The dose rates from other radioactive substances will be about half that from radon. They may be calculated from the following general formula for the dose rate at 1 cm. from a point source of 1 millicurie:

$$\text{dose rate} = 1.10\,\Delta \text{ r.e.p./min.}$$
$$= 0.99\,\Delta \text{ energy units/min.}$$

where Δ is the mean linear ion density in tissue of the particles emitted by the source.

Radium is sometimes mounted on flat plaques of small area covered by very thin Monel metal as a source of β radiation. Though primarily intended for medical applications, such plaques may be found useful for biological research. The energy distribution among the β rays falling on a thin biological specimen irradiated by such a source will be intermediate between the distributions given by Lea for the irradiation of thin and thick specimens. The β-ray dose rate from such plaques and the absorption of the radiation in tissues has been measured by Neary (56) under a variety of experimental conditions. Typically the dose rate was 100 r. per minute at the surface of a plaque, 1.25 cm. in diameter, containing 6.2 mg. radium covered by a 0.05 mm. Monel metal foil. When the plaque was covered by very thin rubber and placed in contact with an extended surface of water the dose rate at depths of 0.5, 1, 2, and 4 mm. was found to be 70, 49, 25, and 8.5%, respectively, of the surface dose. Variation of dose with depth depends critically on the geometrical arrangement employed since β-ray scattering rather than energy loss is the controlling factor.

2. Protons

If a source of polonium is placed immediately behind a film of paraffin wax about 40 μ in thickness, which is sufficient to stop all the α particles emitted by the polonium, a few protons resulting from headon collisions between an α particle and a hydrogen nucleus will be emitted by the wax. The maximum energy of such protons will be 2.7 m.e.v. If aluminum foil is used instead of wax foil rather more protons are obtained as a result of the disintegration of the aluminum by the polonium α particles. These protons will have all energies up to 7 m.e.v. but slow protons will predominate. It appears that a source consisting of 10 millicuries of polonium on a 5 mm. disc mounted 3 mm. behind a 25 μ aluminum foil gives a proton dose rate of about 60 r.e.p. per day at 1 cm. from the foil surface. With increasing thickness of specimen the dose rate falls roughly exponentially and is reduced to half value by about 30 μ of tissue.

Proton beams of sufficient intensity for most biological experiments can only be obtained with the help of high tension equipment, either directly as a result of the acceleration of hydrogen ions or indirectly by the irradiation of tissues by fast neutrons.

A minimum of 1 million volts energy will generally be needed for the irradiation of biological specimens by protons (cf. Fig. 13). Protons of this energy may be obtained by accelerating hydrogen ions supplied from a low voltage arc or other suitable source, by means of a high voltage generator and allowing them to pass through a very thin window at the end of the evacuated accelerating tube. Exposure to δ rays from the window must be avoided as described in the case of electron beams. The apparatus is considerably more complicated than that required for the acceleration of electrons and will be referred to again in connection with neutron sources. Protons have not hitherto been accelerated in this way to greater than 4–5 m.e.v., but 10 m.e.v. may be achieved. Proton beams capable of penetrating about 3 cm. of tissue may be generated by means of cyclotrons of the ordinary type operated by continuous wave high frequency supply. It should be possible to obtain proton beams having a penetrating power of the order of 12 cm. by means of cyclotrons having pole diameters of 90 inches or more operated on the "frequency-modulated" principle (57,58). Wilson (59) has pointed out that the sharp rise in ionizing power of protons during the last few millimeters of their range might be applied to advantage in the radio-

therapeutic use of this radiation. Applying the formula given in
Section C1, it will be seen that only extremely small proton currents
are needed in order to deliver doses of a few thousand roentgens.
If, in a 2–4 m.e.v. proton accelerator, 20 μa. of protons were focused
onto an area of 1 cm.^{2}, and 10% of these emerged through a thin
window, the remainder being absorbed in the supporting grid, the
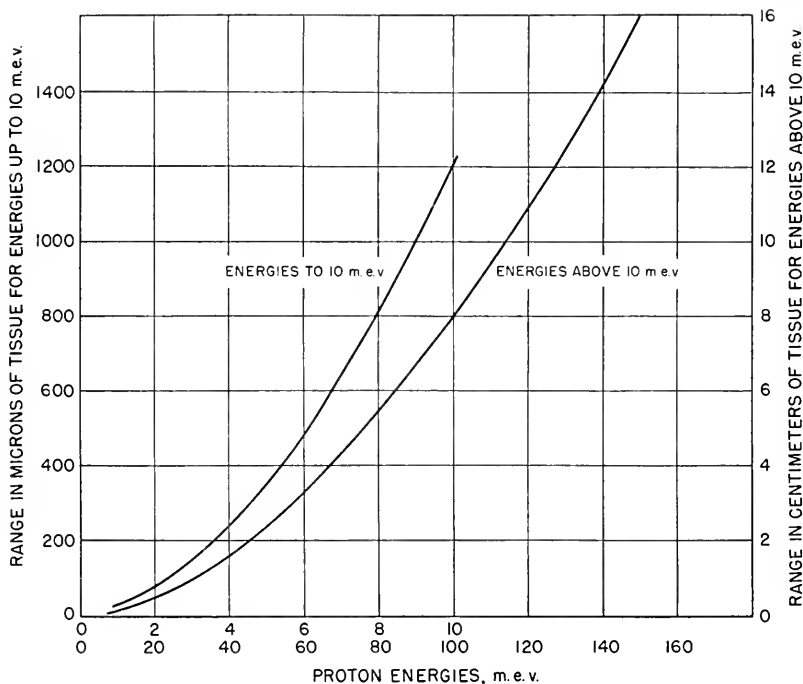
Fig. 13. Curves showing ranges in tissue of protons having
energies 0–10 and 0–150 m.e.v.

mean dose rate beneath the window would be 2 to 3 $\times$ 10^8 r. per
second. The total penetrating power of the protons would be 70–
230 μ of tissue according to the energy, some of which would, of
course, be lost in the window. Ordinary cyclotrons commonly oper-
ate with mean proton currents of the order of 10 μa. and much larger
currents may be obtained. The mode of operation of the cyclotron
is such that the radiation varies somewhat in intensity with the
periodicity of the radiofrequency supply. This intensity variation is

greatly exaggerated in the frequency-modulated cyclotron, in which
the radiation is delivered in pulses, probably about a microsecond in
duration, and repeated 250 times per second. The mean currents
obtained with such instruments will probably not exceed 1% of those
obtained with ordinary cyclotrons.

3. Fast Neutrons

Two years before the identification of the neutron by Chadwick
(60), Bothe and Becker (61) had investigated some of the properties
of neutron radiation resulting from the disintegration of beryllium
by the α particles of polonium under the mistaken impression that it
was a very high energy γ radiation. This original method still pro-
vides the strongest source of neutrons available to laboratories not
equipped with high voltage apparatus, but for biophysical investiga-
tions it suffers from the twofold disadvantage that the total neutron
emission is low and that the neutrons are accompanied by γ rays.
The neutron ionization produced in tissue at 1 cm. from a source con-
sisting of 10 millicuries of polonium mixed with beryllium would only
be about the same as results from exposure to 12 r. per day of γ radia-
tion. There would, moreover, be the problem of filtering out the
γ radiation resulting from the disintegration. For physical investi-
gations radium or radon may be substituted for polonium as the α-
ray source, thereby increasing the neutron emission about fivefold,
but this is inadmissible for radiobiological experiments on account of
the strong γ-ray emission from RaB and RaC in equilibrium with
radium or radon.

Except in rare circumstances, therefore, neutrons for radiobio-
logical research must be obtained from nuclear reactions in which the
bombarding particle is accelerated by some form of high voltage
equipment. The following nuclear reactions may be utilized:

$$D^2 + D^2 \longrightarrow He^3 + n^1 + 3.18 \text{ m.e.v.} \tag{1}$$

$$D^2 + Li^7 \longrightarrow He^4 + He^4 + n^1 + 15 \text{ m.e.v.}$$

$$\text{or less frequently} \longrightarrow Be^8 + n^1 + 14.9 \text{ m.e.v.} \tag{2}$$

$$D^2 + Be^9 \longrightarrow B^{10} + n^1 + 4.2 \text{ m.e.v.} \tag{3}$$

$$D^2 + C^{12} \longrightarrow N^{13} + n^1 - 0.28 \text{ m.e.v.} \tag{4}$$
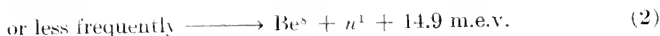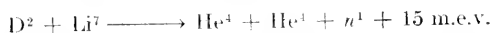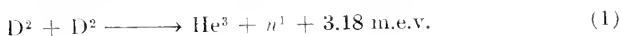
Reaction (1) above has been used by Gray, Read, and collaborators
for a variety of radiobiological investigations. The deuterium ions

were accelerated by means of a 380 kv. constant potential generator
(62). The measured total neutron emission (63) was 10^9 neutrons
per second per milliampere of 380 kv. deuterium ions incident on a
target of "heavy wax." The apparatus was usually operated at ion
currents of 0.5 ma. giving 2 r.e.p. per minute at 1.5 cm. from the tar-
get (r.e.p. unit is defined in Sect. D1). No γ radiation is emitted
in this nuclear reaction, and all the neutrons emitted in a given direc-
tion have the same energy. At the voltages used by Gray and Read
this energy is only slightly dependent on the accelerating voltage since
most of the energy with which the neutron is expelled is derived from
the energy released in the reaction listed to the right of the equation.
The recoil protons generated in tissue by these neutrons have a mean
energy of about 1.2 m.e.v. and a mean linear ion density of 1000 ions
per micron (cf. Table I). which is greater than the corresponding figure
for the other neutrons used hitherto; on this account the biological
results obtained have been more sharply distinguished from those
produced by X rays.

On account of the sharp fall in neutron emission with decreasing
ion-accelerating potentials a minimum of about 300 kv. is required to
give useful dose rates for most biological work. As the accelerating
kilovoltage is increased there is a slow steady rise in neutron output
(Fig. 14) but above 800 kv. this is almost entirely due to the increased
penetration of the deuterium ions into the wax target and may be
estimated from the known relation between range and energy of the
deuteron.

The yield of neutrons from the other reactions listed rises much
more steeply with voltage (cf. Fig. 14). The (D–Li) and (D–Be) re-
actions have both been used as a source of neutrons for biological re-
search (11,24,64).* The yield from both reactions is about the same
as from the (D–D) reaction at 700 kv. and exceeds it at all higher
voltages. There is little accurate information regarding the yields
at voltages above 1 m.e.v. but the curves of Figure 14 may serve as
a guide in the choice of reaction and voltage best suited to particular
needs. The apparatus used by Lea, Haines, and Bretscher (24) was
such that biological specimens could most conveniently be exposed
to neutrons emitted at $90°$ to the deuteron beam. When lithium
was bombarded by 100 μa. of 900 kv. deuterons the dose rate at the
specimens 4 cm. from the center of the target was about 15 equivalent

* The earliest investigations of biological effects of neutrons were made with
the (D–Be) neutrons generated in the 37-in. Berkeley cyclotron.

roentgens per minute. The difficulty of preventing the evaporation of the lithium under intense ion bombardment is a factor that has generally weighed in favor of beryllium targets at the highest energies. The distribution of initial energy among the recoil protons generated by 900 e.kv. deuterium ions bombarding lithium and for 8 m.e.v. deuterium ions bombarding beryllium has been calculated by Lea (*1*, p. 21). Though the energy spectrum extends to 15 and
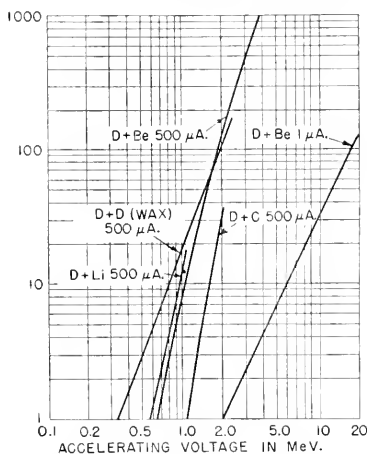
Fig. 14. Curves giving neutron output in equivalent roentgens per minute at 5 cm. from the target for the (D–D), (Li–D), (Be–D), and (C–D) reactions over the voltage range 300 kv. to 5 m.e.v. Data from Reddemann (*96*) for (D–D) neutrons and from Amaldi, Hafstad, and Tuve (*97*) for (Li–D), (Be–D), and (C–D) neutrons. The 500 μamp. curves are for 10% $D^{+}{}_{1}$ and 90% $D^{+}{}_{2}$; the 1 μamp. curve is for 100% $D^{+}{}_{1}$.* Ordinate: dose rate in r.e.p./min. at 5 cm.

to 12 m.e.v., respectively, there is so great a preponderance of low energy neutrons in each case that the mean values are about 4 and 6 m.e.v. compared to about 2 m.e.v. for (D–D) neutrons. The mean initial energies of the particles by which a given cell is ionized are about half these figures. The (D–Be) neutrons are accompanied by γ radiation to the extent of one or more quanta per neutron. Aebersold (*65*) found that the ionization produced in a small amber ionization chamber by the γ rays from beryllium bombarded by 6 m.e.v. deuterons was about 15% of that due to the neutrons. This figure may be taken to indicate roughly the ratio of γ ray to neutron ionization in tissue exposed to the unfiltered beam. Filtration by 3 cm.

* I am indebted to Mr. P. H. Flanders for computing the curves shown in this figure.

of lead effectively eliminated the γ rays from the source. In the article referred to, Aebersold gives very useful practical information concerning the manner in which the (D–Be) neutrons from a 37 inch cyclotron may be collimated so as to give well defined neutron beams suitable for the exposure of parts of large animals, or for therapy. This requires essentially that a wall of wax or tanks of water at least 60 cm. thick be erected close to the cyclotron target to slow down and absorb all neutrons except those passing through the defining channel. This channel has to be lined, preferably with lead, to absorb the 2.2 m.e.v. γ rays resulting from the combination of slow neutrons with protons taking place in the wax or water. These γ rays are in fact more troublesome than those coming from the target (see above). With an arrangement that defined a neutron beam 7×7 cm. at 70 cm. from the target the γ-ray ionization in an amber chamber of 1.75 cc. volume was about the same within and outside the neutron beam, and was about 5% of the neutron ionization within the beam.   The neutron ionization fell to 3, 2, and 1%, respectively, at 5, 10, and 20 cm. from the edge of the beam.   When a second arrangement was used that permitted a neutron beam in a direction of the deuteron beam, instead of at 30° as previously, to be employed, the neutron dose rate in the beam was doubled, but fell off rather more slowly outside the beam.   The γ ray dose rate was unaltered and so was only 2.5% of the neutron dose rate within the beam.

The ionization in the amber chamber at the outlet, which was 70 cm. from the target, was observed to be 5.4 electrostatic units per cubic centimeter per microampere hour of 6 m.e.v. deuterons.   Since the ordinary operating current is stated to be 60 μa. the normal values of the dose rate at 70 cm. may be reckoned as 5 equivalent roentgens per minute.

In 1942 Stone and Larkin (66) reported upon the treatment of 120 cancer patients by neutrons from the 60 inch Berkeley cyclotron. The collimation system was similar to that described by Aebersold. The energy of the deuterons was 16 m.e.v. and the neutron energy spectrum extended up to 21 m.e.v.   The neutron dose rate in the beam at 100 cm. from the target was about 12.5 equivalent roentgens per minute.   The increase in deuteron energy from 6–16 m.e.v. made possible by the use of the larger cyclotron, thus increased the actual yield of neutrons under normal operating conditions about fivefold, and considerably improved the penetrating power of the neutrons so that deep lesions could be more easily reached.   At the same

time, however, the problem of protecting personnel was increased. Moreover, the mean energy of the neutrons, and therefore of the recoil protons that ionize the tissue, has increased so that the mean linear ion density of the ionizing particles has decreased, and is therefore less sharply contrasted with that for the 200 kv. X rays used in therapy (*cf.* Table I).

Microorganisms that do not require a collimated beam could obviously be exposed to neutron dose rates of the order of 10,000 equivalent roentgens per minute if placed 3 cm. from the target of a 60 inch cyclotron. The whole body exposure of animals is lethal at relatively low doses. For mice the acute lethal dose of fast neutrons has been found by Zirkle (*67*) to be of the order of 250 equivalent roentgens and such exposures only yield a limited amount of information. The exposure of tumors or individual organs in small animals requires a collimated beam and presents a new problem. Aebersold gives no data for beams of less than 7 cm. in cross section. Without using collimation, Gray and Read (*68*) exposed tumors implanted in the thighs of mice to neutron doses equal to four times the mean body dose by placing the tumor very close to the target of their (D–D) generator, so that by comparison with the tumor the rest of the body was in a relatively weak radiation field.

1. Slow and Thermal Neutrons

When a neutron collides with a hydrogen nucleus energy and momentum are conserved and any amount up to the full energy of the neutron may be transferred to the hydrogen. It has been shown experimentally by Bonner (*69*) for (D–D) neutrons, in agreement with theory, that the neutron loses on an average half its energy at each collision so that the average speed of a beam of neutrons rapidly falls as the beam passes into a hydrogenous medium. In the case of water or wax most neutrons are ultimately reduced in speed until their energy is equal to the mean energy of thermal agitation of the medium. They are then referred to as "thermal neutrons." As they approach this condition, however, the probability rapidly increases that at any given collision the neutron and hydrogen nucleus will react to form deuterium with emission of a 2.2 m.e.v. γ ray. This is the mechanism by which water and wax shields offer protection against fast neutrons. If it is desired to obtain high fluxes of thermal neutrons, however, it is better to use heavy water or graphite, since,

although these nuclei absorb less energy from the neutron at each collision, they do not react to any appreciable extent with thermal neutrons. For this reason they are used as "moderators" in nuclear fission piles. The recommended way of providing beams of thermal neutrons for biological experiments is therefore to allow a beam of fast neutrons to pass into a mass of very pure graphite or heavy water. Prior to the construction of "piles" only two biological experiments with slow neutrons were reported (70,71). In each case the fast neutrons from a cyclotron were slowed down by passage through wax or water. Specimens impregnated with lithium or boron were exposed with the object of looking for a biological effect due to ionizing particles resulting from the disintegrations:

$$\text{Li}^6 + n^1 \text{ (slow)} \longrightarrow \text{H}^3 + \text{He}^4 + 4.5 \text{ m.e.v.} \tag{5}$$

$$\text{B}^{10} + n^1 \text{ (slow)} \longrightarrow \text{Li}^7 + \text{He}^4 + 2.5 \text{ m.e.v. (see ??)} \tag{6}$$
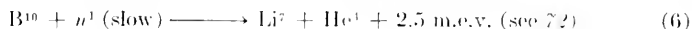
Thermal neutrons will in general be accompanied by rather more energetic neutrons. Fortunately, however, the thermal neutrons are strongly and selectively absorbed in cadmium so that an effect of the thermal neutron alone can be looked for as a difference. It seems that, at least in the experiments of Kruger, a biological effect of the disintegration particles was established.

Very little indeed has been published concerning the biological effects of thermal neutrons on untreated tissues. Practically the whole of the ionization produced in biological specimens a millimeter or so in thickness exposed to pure thermal neutrons will be produced by protons resulting from the disintegration of the nitrogen content of the tissue:

$$\text{N}^{14} + n^1 \longrightarrow \text{C}^{14} + p + 0.62 \tag{7}$$

When, however, thermal neutrons fall on a layer of tissue a few centimeters in thickness the reaction with hydrogen, referred to above, namely:

$$\text{H}^1 + n^1 \longrightarrow \text{H}^2 + h\nu + 2.2 \tag{8}$$

will generate 2.2 m.e.v. γ rays, which will produce more ionization in the tissue than the disintegration of nitrogen. For many types of biological reaction, particularly the types leading to the gross injury of somatic cells, the proton ionization, being of high linear ion density, will be much more effective than an equal amount of γ-ray ionization. Allowing a factor of ten for this intrinsic biological ef-

fectiveness, Mitchell (73) calculates that when a large animal is exposed to a beam of thermal neutrons disintegration of nitrogen will account for about 60% and the γ rays for about 40% of the total biological effect in the superficial cells.

The nuclear fission pile provides larger fluxes of thermal neutrons than any other source; Zirkle (67) has estimated the acute lethal dose of thermal neutrons for mice by exposing animals in a pile. The animals were introduced through a tunnel in the pile shield to a position at which they would be exposed to a large flux of slow neutrons. The exposure chamber was made of bismuth to absorb γ rays, and bismuth partitions were arranged so that each animal was shielded from the γ rays emitted from its neighbors as a result of the slow neutron capture by hydrogen (reaction 8). Half the irradiated animals survived three weeks after exposure to a slow neutron dose of 4.2 $\times$ 10^{12} slow neutrons per cm.^{2} under the conditions of this experiment. Other radiations were present and it is interesting to note that the contributions to the total ionization in the tissues were estimated to be: fast neutrons 45 "n" (units) (defined in Sect. D4); γ rays 84 r.; protons from the disintegration of nitrogen 54–68 n; radioactivity induced in the tissues by neutron bombardment 10 r. Since previous work had shown that as regards acute lethal effects in mice 1 n of neutrons was about as effective as 9 r. of β or γ radiation, the four contributions to the total effect of exposure in the pile were assessed as equal or equivalent to 45 n, 10 n, 54–68 n, and 1 n, respectively, totaling 109–123 n, in fair agreement with the measured acute mean lethal dose for fast neutrons, which had been found to be 91 n. The conclusion was drawn that the results exclude the possibility and the fear of the existence of some unknown mechanism that creates a hazard far greater than that due to the known mechanisms. It will be seen that the contribution of slow neutrons to the total effect was only estimated to be 50% and, judging by the figures given for the slow neutron flux, may barely have exceeded 30%.

These figures serve to emphasize the extreme technical difficulty that at present attends an investigation of the biological effects of thermal neutrons obtained by the passage of fast neutrons through a moderator. It is possible that slow neutrons could more easily be obtained from the (D–C) reaction (reaction 4, Sect. C3). This reaction has not hitherto been used for biological work. The yield is much less than from the (D–D) reaction even at 1 m.e.v. and the product nucleus, N^{13}, is radioactive with a half-life of ten minutes.

The decay is by positron emission and the positrons by annihilation give 0.5 m.e.v. γ radiation. A nuclear γ ray is also emitted. By the end of relatively short bombardments, therefore, each neutron will be accompanied by more than 1 m.e.v. of γ radiation. Despite these disadvantages neutrons from this reaction may have a place in biological research as providing neutrons of uniform but low energy. For example, when the accelerating voltage is 800 kv. the neutrons in the forward direction have about 350 kv. energy (74); at 600 kv. and 90° to the beam the neutrons have about 150 kv. energy. The mean energy of the recoil protons will be half these values, which, as pointed out in earlier sections, is an interesting energy region since the protons are too slow to form any δ rays and the ratio of excitation to ionization may be different from that characteristic of faster particles.

5. Alpha Particles

Polonium with a half-life of 136 days emits α particles having a range of 4 cm. in air and about 40 μ in tissue unaccompanied by β or γ radiation. A small disc on which polonium is deposited is therefore an extremely convenient source of α radiation for the irradiation of bacteria (75), spores (76), and other small objects (cf. 29). All the α particles are emitted with the same energy (5.3 m.e.v.) and all are slowed down to approximately the same extent by foils of mica or other material placed in front of the source. Zirkle (76) took advantage of this fact to study the relative effectiveness of α particles of different speed in inhibiting division in fern spores. The α-ray dose rate at 1 cm. from a point source of 1 millicurie is given by a similar formula to that for β radiation (cf. page 546), namely:

$$\text{dose rate} = 1.20 \ \Delta \text{ r.e.p./min.}$$

Δ = ion density of α particles in ions/μ tissue. Thus for α particles emitted by polonium the dose rate is about 2700 r.e.p./min. rising to 3100, 3800, 5000, and 7500 r.e.p. per minute as the energy is reduced by foils successively to 4, 3, 2, and 1 m.e.v. RaC' sources, prepared by exposing a metal disc to radon and allowing half an hour to elapse for the decay of RaA, emit α particles of 7.68 m.e.v. energy. Although β and γ rays will also be emitted from the source prepared in this way the ionization per unit volume due to the β radiation in any specimens small enough to be traversed by the α particle will be only about 1% of the α-ray ionization and the effect of the γ radiation will

be negligible.   The millicurie activity of such sources after decay of
RaA will be perhaps a quarter that of the radon used in their prepara-
tion, and the activity decays with a half-life of somewhat less than
half an hour (29).

The exposure to α radiation of specimens larger than the α-ray
range may sometimes be accomplished by immersion in a solution
containing radon.   This method has been used for the irradiation of
root tips (77), microspores contained within the intact anther (2),
and tadpoles (78).   Radon solution has also sometimes been used for
the irradiation of viruses (79) and bacteria (80, Chapter V) in sus-
pension, and in a variety of radiochemical studies.   When a specimen
is uniformly perfused by radon in equilibrium with its decay products
RaA and RaC′, which emit α rays, the α-ray dose rate is 810 C
r.e.p. per minute, where C is the concentration of radon in milli-
curies per gram of tissue.   If required, therefore, many thousands
of roentgens can conveniently be delivered in this way.   The speci-
mens will be exposed in addition to β radiation from RaB and RaC
to an extent that depends on the volume of the radon solution in
which the specimen is immersed.   At most the β-ray dose will be
5% of the α-ray dose.   The γ-ray dose will generally be negligible.

Application of this technique to the irradiation of roots was in-
vestigated by Gray and Read (77) in some detail.   They found that
when the root was introduced into the solution the radon entered at
roughly the rate to be expected for simple diffusion of an inert gas of
high molecular weight, and at equilibrium the concentration was the
same in the root as in the solution.   When the root is transferred
again to pure water the radon diffuses out at the same rate.   In each
case half equilibrium concentration is reached in the root tip in about
1.5 minutes.   The atoms of RaA and RaC′ formed in the solution do
not enter the root and those formed in the root do not leave.   In these
circumstances, the total dose is related to the radon concentration in
the root in the manner given above even when the exposure is too
short for radioactive equilibrium to be established between radon,
RaA, and RaC′, provided the radon concentration reaches its equilib-
rium value at the center of the root.   The dose is, of course, spread
over a longer period than the duration of immersion because it is de-
livered in part by the RaA and RaC′ remaining in the root.   The
RaA and RaC′ formed in the solution were found to accumulate on
the surface of the root thereby very greatly increasing the dose to the
superficial cells.   This effect can give rise to entirely misleading re-

sults when small organisms having a large ratio of surface to volume are irradiated. It may be minimized by dissolving the radon in 2% gelatin solution instead of in water. The RaA and RaC′ atoms apparently then become attached to protein molecules and are much less mobile. In the experiments of Kotval and Gray (2) shoots bearing inflorescences, stripped of their bracts, sepals, and petals were immersed in radon solution. Even with the modification of dissolving the radon in 2% gelatin the absorbed activity on the surface of the anthers increased the mean dose to the microspores by 10%.

When tadpoles were allowed to swim freely in radon solution the attainment of an equilibrium concentration throughout the animal was aided by imbibition and a blood circulation, so that the time taken to reach half equilibrium concentration was not much greater than in the case of roots. If diffusion alone were operative the time would be proportional to the square of the linear dimensions. In the case of larger animals the uniformity of radon distribution should be carefully investigated. Radon is very much more soluble in many organic substances than in water. Radioautographs showed slight evidence of inhomogeneity in the distribution of radon in the tadpole. Studies on radioactive xenon (99), which probably closely resembles radon as regards solubility, have shown very large variations indeed in concentration in different organs of the rat. In view, however, of the importance of comparing wherever possible the efficiency of high and low ion density radiation it is worth while to take some trouble in the design of an experiment so as to permit the use of α radiation, more especially in laboratories not provided with neutron sources.

D. MEASUREMENT OF IONIZING RADIATION

In the preceding sections numerous instances have been given of the way in which the same total dissipation of energy in tissue by ionizing radiations may affect vital processes to a different extent according to the distribution of ions within the cell, the duration of irradiation, and other factors. These differences were only brought to light after means had been devised for estimating the total energy dissipated by the ionizing particles in all forms—ionization, excitation, chemical change, and heat—per unit mass of tissue.

The quantity of energy with which we are concerned is 93 ergs or 2.2×10^{-6} cal. per gram of tissue per roentgen. For technical reasons it would only be possible to attempt to measure this energy calori-

metrically when the dose is large and the duration of irradiation short, and even in such cases it would be unwise to attempt to measure the rise in temperature of the biological specimen itself since this might be seriously disturbed by an alteration in cellular metabolism initiated by the radiation.   The rise in temperature of an object representing the biological specimen in size and composition might be measured, but here again care must be exercised with regard to thermochemical change.

1. Units of Dose

The Roentgen.   Since it is impossible to measure small doses calorimetrically alternatives have to be sought.   The problem first became urgent more than 25 years ago in connection with the measurement of X-ray dose in radiotherapy and the most practicable solution was found to be to attempt to measure the ionization produced by the X radiation in air under conditions that made it possible to infer the amount of X-ray energy that would be absorbed per unit mass of air in the position occupied by the portion of tissue under consideration.   The quantity measured is therefore the ionization produced in air under standard conditions, and the unit of X-ray dose, the roentgen, is defined as "that quantity of X or γ radiation such that the associated corpuscular emission per 0.001293 gm. of air produces, in air, ions carrying 1 electrostatic unit of quantity of electricity of either sign."

The usefulness of the roentgen as a unit of X- or γ-ray dose depends on three things.   First, that the total number of pairs of ions formed by X rays per unit mass of air bears an approximately constant ratio to the *total* X-ray energy absorbed per unit mass of air over practically the whole range of X and γ radiation.   Secondly, that at least to a first approximation the absorption of X and γ rays per unit mass of air is equal to the absorption per unit mass of tissue, and even a rough knowledge of the quality of the X radiation and the composition of the tissue enables the ratio of the two to be estimated with accuracy.   Thirdly, the range of the secondary electrons that actually ionize the tissue is only a few millimeters even when generated by γ radiation so that the ionization is produced in tissue at a point not far removed from that at which X- or γ-ray energy is absorbed.   The pattern of actual energy absorption throughout the bulk of tissue is thus not greatly different from the pattern of the distribution of dose measured in roentgens.

This state of affairs is, of course, not even approximately true in the case of megavoltage X radiation, and future work with this radiation derived from betatrons, synchrotrons, and microwave linear accelerators will call for a reconsideration of the manner in which dose can be most usefully defined. This problem has only just become acute.

Much earlier the use of neutrons in research and in therapy called for a reconsideration of the problem of radiation dosimetry, on account of the fact that in the case of neutron radiation the second condition upon which the usefulness of the roentgen hinges is completely invalid. The absorption of neutron radiation per unit mass of tissue is generally about five times that per unit mass of air and depends critically on the hydrogen content of the tissue.

More recently, in connection with the use of radioactive isotopes in research and therapy, problems have arisen connected with the estimation of the dose received by tissues on account of their radioactive content. Although such cases may be covered by an obvious extension of the definition of the roentgen, this unit cannot properly be applied as at present defined.

Analyzing the present situation we see that the immense practical value of the roentgen is that it specifies a quantity of energy too small to be measured calorimetrically by the ionization that would be produced in air by the expenditure of this amount of energy by ionizing particles; its present inadequacy arises from the fact that the quantity defined is X- or γ-ray energy absorbed in air, whereas the quantity in which we are interested is the energy expended by the ionizing particles in tissue. There can hardly be any doubt that the wise course is to define dose in terms of the energy lost per unit mass of tissue by the ionizing particles that traverse the element of tissue under consideration. Dose may then be expressed in the fundamental units of physics, namely ergs per gram, for all types of ionizing radiation. This introduces a great simplification into our conception of dose. If it is desired to maintain a numerical correspondence with dose expressed in roentgens, a new unit must be defined which is a multiple of the natural and fundamental unit of 1 erg per gram. This thought prompted the definition of the "energy unit" of dose by Gray and Read (81) and of the "equivalent roentgen" (e.r.) or "roentgen-equivalent-physical" (r.e.p.) which is attributed to Parker (82). The energy unit and the equivalent roentgen only differ in the quantity selected as the unit of energy.

The Energy Unit. The unit proposed by Gray and Read adopts the amount of energy lost by the ionizing particles per unit mass of water exposed to 1 r. of γ radiation, so that on the basis of currently accepted physical data:

$$1 \text{ energy unit } = 93 \text{ ergs/g. tissue}$$

This unit has the advantage that the dose received by tissues exposed to hard X rays and γ rays has the same numerical value whether expressed in roentgens or energy units, so that if the unit gains general acceptance the doses employed in earlier X-ray and γ-ray work, expressed in roentgens, would be strictly comparable, without numerical adjustment, with later doses expressed in energy units.

The Roentgen-Equivalent-Physical. The roentgen-equivalent-physical adopts as the unit of energy the energy lost by fast electrons in producing in air ions of either sign carrying 1 electrostatic unit of charge, so that the accepted data lead to:

$$1 \text{ r.e.p. } = 84 \text{ ergs/g. tissue}$$

Tissue exposed to 1 r. of γ radiation receives a dose of 1 energy unit or 1.1 r.e.p.

2. Measurement of Ionization

Figure 15a shows diagrammatically one method of measuring ionization currents, which has proved convenient (83). The passage of ionizing particles across the air volume defined by the ionization chamber gives rise to the formation of pairs of ions that, in the absence of an electric field, rapidly recombine. If the chamber wall is charged to, say, 100 v. positive with respect to the collecting electrode, which is at earth potential, the positively charged ions move toward the collecting electrode and the negative ions toward the walls; and if the electric field is sufficiently large virtually all the ions of one sign will be collected by the inner electrode and none will be lost by recombination with ions of opposite sign. It is, of course, of cardinal importance to an absolute measurement of dose that no appreciable fraction of the ions should be lost. The ease with which this condition can be fulfilled varies with the nature of the gas being ionized, the gas pressure, and the type of ionizing particle. Difficulties are greater at high pressures than at atmospheric pressure, and increase with increasing linear ion density, i.e., generally with particle type in

the order β rays, protons, α rays. Much detailed information on this subject is available (*30,84*).

The collecting electrode is connected with the measuring instrument by a highly insulated wire traveling inside an earthed tube. If the dose rate to be measured is 1 r. per minute and the volume of the ionization chamber is 1 cc., the current to be measured is roughly 5×10^{-12} amp. The current-measuring instrument must therefore be some form of electrometer (*85*, Ch. IV), electrometer valve, or similar device (*86,87*), and special attention must be given to the quality of the electrical insulation.
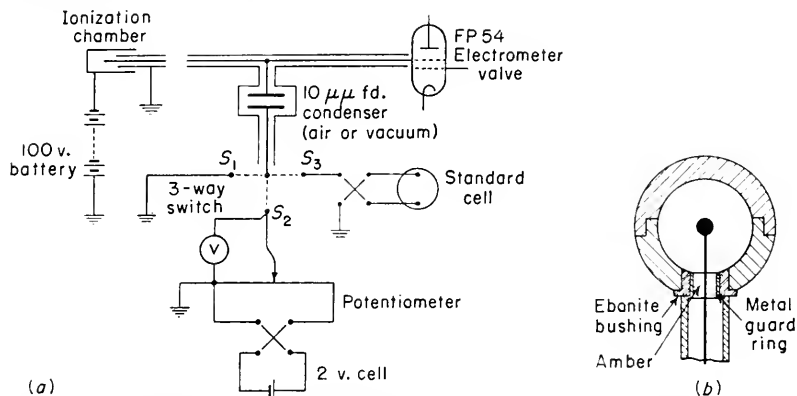
Fig. 15. (*a*) Diagram of Townsend balance method of measuring ionization current. (*b*) Representative diagram of an ionization chamber.

The diagram shows the collecting electrode connected to the control grid of an FP 54 electrometer valve. It is possible to measure the current by noting the rate at which the potential of the grid rises, as judged by the change in anode current. There are several advantages, however, irrespective of the nature of the measuring instrument, in keeping the collecting electrode near earth potential by some system that balances out the ionization current. The method illustrated in Figure 15a accomplishes this by gradually changing the potential of one plate of the fixed condenser (*83*). If such a measuring system is employed there should exist a negligible difference of potential between the collecting electrode and the earthed cylinder which enters the base of the ionization chamber, thus imposing a minimum electrical strain on the insulator which supports the collecting electrode. The critical insulator is preferably made of amber, not only because amber is an excellent insulator but because its insulating

properties are less affected by ionizing radiations than those of most
other substances.   The insulation between the earthed cylinder and
the chamber wall is less critical.   The actual ionization chamber
should comprise the components shown in Figure 15*b*, namely, an
envelope to define the volume of air ionized (the inner surface of
which at least must be conducting), insulation, an earthed cylinder,
high grade insulation, and the central wire supporting the collecting
electrode.   The shape and material of the wall and electrode will be
varied according to the type of ionization to be measured.

The apparatus described is suitable for the measurement of dose
rate.   If it is desired to measure total dose similar apparatus may be
employed but with an integrating circuit following the electrometer
valve, or the "condenser dosimeter" method may be used (*85*, Ch.
IV; *88*).

3. Energy Equivalent of One Electrostatic Unit of Ions Formed in Air

Denoting the average energy lost by an ionizing particle in pro-
ducing a pair of ions in the gas within the chamber by W, the total
energy lost by the ionizing particles per unit mass of gas $(E_m)_{gas}$ is
given by:

$$(E_m)_{gas} = \frac{W \times J_v}{(\text{density})_{gas}} \tag{9}$$

where J_v is the number of pairs of ions formed per unit volume.

The quantity W has been measured experimentally under various
conditions for electrons, protons, and α particles.   It is nearly, but
not quite independent of particle type and particle speed in the case
of air (*cf.* Table V).   It is different for different gases, and is not
necessarily an additive property in the case of gas mixtures, so that
the value of W must be ascertained if gas mixtures not previously in-
vestigated are used in the chamber.

Reviewing the literature, Gray (*23*) concluded that there are good
grounds for believing that the average energy expended by electrons
in producing a pair of ions in air is independent of electron energy
above about 300 kv. and is equal to 32.0 e.v.   Below 300 kv. there
was a steady rise in W with decreasing electron energy, which is
slight at first and steeper at very low energies until at 1 kv. the value
is probably about 37 e.v.   Neglecting finer details, it would seem that
the general mean value $(W_e)_{air} = 32.5$ e.v. or 5.17×10^{-11} erg may

be adopted for most purposes as representing the energy expended per ion pair formed by electrons. Consistent with the rise in W_e with decreasing electron speed is the fact that the value of W for the more slowly moving protons and α particles is greater than that for electrons. The best experimental values appear to be $(W_p)_{\text{air}} = 36$ e.v. for protons of energy between 2 and 8 m.e.v. and $(W_\alpha)_{\text{air}} = 35.1$ e.v. as the mean over the whole range of RaC′ α particles, and $(W_\alpha)_{\text{air}} = 35.6$ e.v. for polonium α particles.

4. Design of Ionization Chambers and Choice of Conditions of Irradiation for Experiments with Particular Sources

External Beta-Ray Sources. Thought given to the choice of the conditions of irradiation will often greatly facilitate the accurate measurement of the dose received by the specimens. For example, when using β radiation it is necessary to bear in mind not only the ease with which slow β particles are scattered through large angles but also the fact that the extent of the scattering varies with the atomic number of the scattering medium. If specimens such as bacteria are irradiated in thin films of gelatin, the dose is probably most accurately estimated by means of a very shallow ionization chamber having top and bottom walls made out of the thinnest obtainable paper or nylon, rendered conducting by graphite, and an open grid collecting electrode. The chamber is situated beneath a stop, as in Figure 16a. The stop is made of low atomic number elements and defines the β-ray beam. By using stops of different diameter, reflection from the edges can be allowed for. Only the depth of the chamber needs to be accurately defined.

If I is the total measured ionization in electrostatic units per minute, R the radius of the upper surface of the stop, d the perpendicular distance of this surface from the source, t the depth of the chamber, the dose rate to the specimen placed on the level of the upper surface of the stop is given by:

$$\frac{kI}{2\pi R^2 t \left[\dfrac{d^2}{R^2} \left(1 - \dfrac{1}{[1 + (R^2/d^2)]^{1/2}} \right) \right]} \tag{10}$$

which, when $R/d \ll 1$, reduces to:

$$kI/\pi R^2 t \tag{11}$$

where k is the ratio of the mass stopping powers of tissue and air for the electrons in question if the dose is expressed in r.e.p. per minute and the ratio of the electronic stopping powers if doses are expressed in energy units per minute.
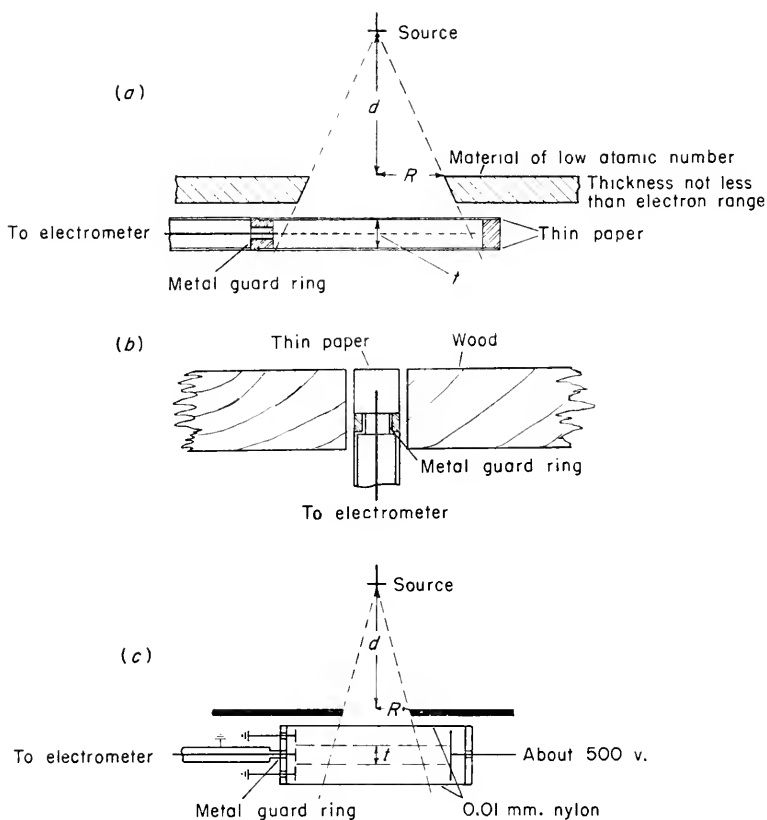
Fig. 16. Arrangement of ionization chambers for measuring (*a*) the dose of β rays when the biological material is irradiated in a thin film, (*b*) the dose of β rays to the superficial cells of a portion of tissue, and (*c*) the dose of α rays delivered to unicellular organisms.

If the specimens are more conveniently irradiated on the surface of agar, or if we are interested in the superficial layer of cells of a mass of tissue, a small cylindrical chamber may be arranged somewhat as shown in Figure 16*b* to measure the dose. Lateral scattering, which will occur in the irradiation of the specimen, is provided by wood. If

the agar is contained in a metal dish, the thickness of agar should be comparable with the maximum β-ray range so that there is no ambiguity as to the nature of the scattering medium.

The dose rate with this arrangement is simply J_v r.e.p. per minute where J_v is the ionization per unit volume at $0°C$. and 760 mm. of mercury. Since it is difficult to know the volume of a paper chamber accurately it is generally more convenient to calibrate the chamber by covering it with a 4 mm. thick graphite cap and to measure the γ-ray ionization produced by a radium source of known strength. It may be assumed that the ionization produced under such circumstances due to the γ rays of radium filtered through 0.5 mm. platinum is at the rate 8.3 electrostatic units per cubic centimeter per hour at 1 cm. from a point source of 1 mg. For further details, several recent papers should be consulted (56,89,100).

External Proton and Alpha-Ray Sources. Scattering is negligible in the case of protons and α particles but a technical problem arises in connection with the collection of the ions, particularly in the case of α-ray ionization. A field of 500 v. per centimeter fails to collect about 20% of the ions formed in air by α rays traveling in the direction of the electric field (84,90). The same field collects all but about 2% of the ions when the particles are traveling perpendicular to the field. It is therefore advisable always to define the beam to be used by a stop, as in the first method described for β rays, but to use a transverse collecting field between parallel plates, which define the depth of the volume from which ions are collected, as, for example, in the manner shown in Figure 16c. The dose rate to specimens at the level of the upper surface of the stop will be given by the same formula as for the β-ray dose rate but with the inclusion of a factor f to allow for the different value of W_{air} for protons and α particles. Thus:

$$\text{dose rate} = \frac{fk'I}{2\pi R^2 t \left[\dfrac{d^2}{R^2} \left(1 - \dfrac{1}{[1 + (R^2/d^2)]^{1/2}} \right) \right]} \tag{12}$$

or, when R/d is small:

$$fk'I/\pi R^2 t \tag{13}$$

where $f = 1.11$ for protons and 1.09 for α particles; and k' is the mass stopping power ratio of tissue and air for either protons or α particles if the dose is expressed in r.e.p. per minute or the electronic

stopping power ratio if the dose is expressed in energy units per minute.

If the source is quite clean and not covered by any kind of film so that the energy of the α particles is accurately known, it may be easier to count the number of α particles emitted within a known small solid angle and apply the general equation given on p. 525:

$$\text{dose rate at 1 cm.} = 1.10 f \Delta \text{ r.e.p./min.} \qquad (14)$$

$$= 0.99 f \Delta \text{ energy units/min.}$$

for a point-source of 1 millicurie strength, *i.e.*, a source emitting $3.7 \times 10^7 \alpha$ particles isotropically, where Δ is the ion density of the particles in water and f as above is 1.09.

Nonionizing Primary Radiation. *Fast Neutrons.* The difficult problem that arises by virtue of the fact that the absorption of neutron energy per unit mass of air bears little relation to the absorption in tissue was met in America by recording always the reading of a particular commercial X-ray ionization chamber, the Victoreen 100 r. chamber, when exposed to the neutron beam, but expressing the fact that the ionization was produced by neutrons and not by X rays by substituting the symbol n for the symbol r. This was recognized as a temporary expedient that might at least enable particular conditions of exposure to be reproduced pending further research on neutron dosimetry. Owing to the fact that the Bakelite walls of the Victoreen ionization chamber contained hydrogen the measured ionization is in part due to recoil protons and bears a closer relation to the absorption of neutron energy in tissue than would the ionization produced in free air. As a result of a very careful subsequent study of the problem, Aebersold and Anslow (*91*) concluded that under most of the conditions that obtained in the previous biological investigations carried out in the United States with fast neutrons, the energy dissipation in tissue could be reckoned to have been just about double that employed by exposure to 1 r. of X radiation. Thus it was concluded that for fast neutrons:

$$1\ n = 2 \text{ to } 2.5 \text{ r.e.p.} = 168 \text{ to } 210 \text{ ergs/g. tissue}$$

The problem of the measurement of fast neutron dose is that of measuring as accurately as possible and in absolute units, with the aid of ionization chambers, the amount of energy lost by ionizing particles per unit mass of tissue when that tissue is exposed to a beam of

neutron radiation. The problem has been discussed in some detail
by Zimmer (92), Aebersold and Anslow (91), and Gray (23). In es-
sence, the problem is the same as that of measuring the energy ab-
sorbed by tissue exposed to X radiation or which is uniformly per-
meated by a radioactive substance, and the solution depends on the
fact that, if into any solid medium in which ionizing particles are
being uniformly generated a small empty cavity is introduced, the
number and speed of the particles crossing any small area of the cavity
in any given direction will be the same as for the particles crossing an
equal area similarly situated in the solid. This is self-evident in the
case of protons for the particles moving in such a direction as to enter
the cavity, for it is clear that since there is no appreciable scattering
of protons removal of solid material to the right of area A in Figure 17
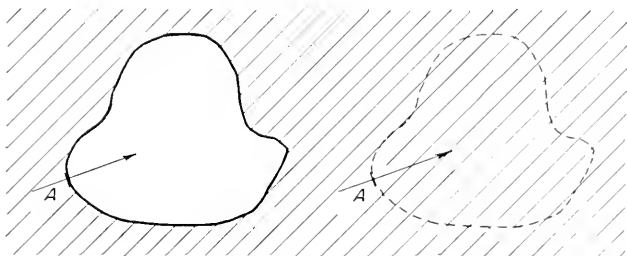
Figure 17

to form the cavity can have no influence on the protons crossing the
boundary into the cavity. It may be shown to be the case, also, for
particles moving in the opposite direction. It is, in fact, true gener-
ally for all types of particle, including particles that are easily scat-
tered through large angles, such as electrons.

Suppose the cavity to be filled with gas. A group of particles
crossing area A in any given direction will lose a certain amount of
energy in the course of ionizing the gas. The same group of particles
passing across an equal and similarly shaped volume of solid (Fig.
17) will lose very much more energy. The ratio of the two energy
losses may be denoted by ρ. This ratio is known as the stopping
power of the solid relative to the gas for the particles in question.
The ratio is to a first approximation equal to the ratio of the densities
of the solid and the gas. The same will be true for all groups of par-
ticles crossing the cavity and the equivalent groups moving across an
identical volume of the solid. Thus, if ρ is taken to denote the mean

value of the stopping power of the solid relative to the gas for all groups, the energy lost by the ionizing particles per unit volume of solid is equal to ρ times the energy lost per unit volume of gas. The energy lost in the gas is equal to the number of ions formed multiplied by the average energy, W_{gas}, expended in the production of a pair of ions in the gas. We have, therefore:

$$(E_v)_{solid} = \rho \times W_{gas} \times (J_v)_{gas} \qquad (15)$$

where E_v is the energy lost by the ionizing particles per unit volume of solid $(J_v)_{gas}$ is the number of ions formed per unit volume of gas and ρ is the linear stopping power of the solid relative to the gas. Alternatively:

$$(E_m)_{solid} = \rho_m \times W_{gas} \times (J_m)_{gas} \qquad (16)$$

where the energy and ionization are expressed per unit mass and ρ_m is the mass stopping power.

If the solid and gas have the same atomic composition, then it is generally assumed (23)* that, irrespective of the state of chemical combination of the atoms, ρ_m is unity, so that:

$$E_m = W_{gas} \times J_m \qquad (17)$$

The ideal method of estimating E_m is therefore to measure J_m in a small ionization chamber of which both walls and gas have the same atomic composition as the tissue to be irradiated. The dose in equivalent roentgens is by definition E_m divided by E_0, the X-ray energy absorbed per unit mass of air exposed to 1 roentgen of X rays. The X-ray energy absorbed in 0.00129 g. of air exposed to 1 roentgen generates, by definition, ions of either sign carrying 1 electrostatic unit of charge. If, therefore, ionization is expressed throughout in electrostatic units of charge collected:

$$E_0 = (1/0.00129) \times 32.5 = 32.5/d. \text{ (e.v./g.)}$$

where d. = density of air at 0°C. and 760 mm. of mercury. Therefore the dose received by any material in equivalent roentgens follows from equation (17); thus:

$$\text{dose} = \frac{E_m}{E_0} = \frac{W_{gas} \times J_m}{E_0} = \frac{W_{gas} \times J_v}{\text{density of gas} \times E_0}$$

and inserting the value of E_0:

* Recent experiments by Appleyard (105), however, reopen the question.

$$\text{dose} = \frac{\text{d}}{\text{density of gas used}} \cdot \frac{W}{32.5} \; J_v \; \text{r.e.p.} \qquad (18)$$

where J_r is the observed ionization in electrostatic units per cubic centimeter. When the ionization chamber is filled by a gas having the same atomic composition as the walls, the only limitation on its size is that the neutron intensity should be sensibly constant throughout the volume. Neutron ionization measurements have been made in chambers containing gases approximating the walls in atomic composition by Gray and Read (*93*) and Aebersold and Anslow (*91*).

In a private communication, Prof. Failla informed the author that H. H. Rossi, working in his laboratory, has found that tissue having the composition H = 9.75%, O = 70.98%, C = 15.71%, and N = 3.55% may conveniently be represented by walls composed of: gelatin, 20.15%; glycerin, 5.18%; water, 66.23%; sugar, 8.43%; and a gas mixture having the following composition (partial pressures in cm. of Hg): methane, 29.22 cm.; hydrogen, 50.12 cm.; oxygen, 48.96 cm.; air, 3.53 cm. The gas mixture, however, is explosive.

It is always more convenient, however, if air can be used in the chamber instead of a special gas filling. If the results are to be readily and quantitatively interpreted, two conditions must be fulfilled. The dimensions of the chamber and the air pressure must be such that the particles generated in the walls lose a small fraction of their energy in crossing the chamber, and it must be possible to estimate the value of ρ_m, the mass stopping power of the walls relative to air.

A knowledge of the quality of the neutron radiation being measured and reference to tables of ranges (Table IV and Fig. 13) would generally indicate whether the former condition is likely to be fulfilled, but it can also be tested experimentally, for, if it is fulfilled, the observed ionization will be strictly proportional to pressure for all pressures up to the working pressure.

Many interesting tests of this kind were made by Aebersold and Anslow with a variety of gas pressures and wall materials. Provided the chamber wall is made of a material about as rich in hydrogen as tissue, the condition is well fulfilled for gas volumes of the order of 1 cc. and for neutrons having upward of 2 m.e.v. energy.

The mean stopping power of the wall material, ρ_m, varies with the energy of the recoil protons and therefore with the neutron energy. Data given by Aebersold and Anslow and by Gray (*23*), reproduced in Table IV, enable ρ_m to be estimated with sufficient accuracy for neutrons of energy between 0.1 and 15 m.e.v.

TABLE IV

Energy, Range, and Stopping Power Data Appropriate to Neutrons of Given Energy

Neutron energy, m.e.v.	0.1	0.4	1.0	2.0	4.0	8.0	12.0
Maximum energy of recoil atoms, m.e.v.							
H	0.1	0.4	1.0	2.0	4.0	8.0	12.0
C	0.0284	0.148	0.284	0.568	1.136	2.272	3.408
O	0.0221	0.0884	0.221	0.442	0.884	1.768	2.652
Range of recoil atoms of max. energy, cm. air (760 mm. Hg, 15°C.)							
H	0.138	0.65	2.3	7.2	23.1	77.3	158
C	0.024	0.060	0.11	0.16	0.24	0.36	0.46
O	0.020	0.046	0.074	0.12	0.18	0.26	0.34
Atomic stopping power, relative to air, for recoil particles of half-max. energy							
Recoil protons							
H	0.390	0.316	0.256	0.231	0.214	0.201	0.194
C	(0.98)	(0.96)	0.923	0.903	0.889	0.878	0.872
O	(1.01)	(1.02)	1.010	1.054	1.063	1.070	1.072
Recoil carbons							
H	...	...	...	...	0.39	0.36	0.34
Recoil oxygens							
H	...	...	...	...	0.42	0.38	0.36

Figures in parentheses for stopping power of carbon and oxygen relative to air for protons are obtained by extrapolation, assuming the values approach unity at zero particle energy. The stopping power of carbon relative to air for all energies of carbon and oxygen recoils listed probably lies between 0.97 and 1.0; that of oxygen between 1.02 and 1.0.

The expression for the neutron dose, in equivalent roentgens, delivered to the material of the walls of the chamber or to any material having the same atomic composition as the walls, when the ionization produced in the air in the chamber at 0°C. and 760 mm. of mercury is J_v electrostatic units per cubic centimeter, is given by:

$$\text{dose} = \rho_m \times \frac{(W_p)_{\text{air}}}{32.5} \; J_v \quad \text{r.e.p.} \tag{19}$$

where ρ_m and $(W_p)_{\text{air}}$ are mean values for the particular energy distribution of recoil protons generated by the neutrons.

TABLE V

Computed Variation of W_{air} with Particle Energy for Protons and α Particles

α particle energy, m.e.v.	1	2	3	4	5	6	7	8
Proton energy, m.e.v.	0.25	0.5	0.75	1.0	1.25	1.5	1.75	2.0
W_{air} for an α particle of given *instantaneous* energy.........	36.9	35.8	34.8	34.5	34.3	34.2	34.1	34.0
W_{air} relative to that for an 8 m.e.v. α particle.	1.085	1.053	1.024	1.015	1.009	1.006	1.003	(1.000)
W_{air} for an α particle of given *initial* energy..	38.6	37.3	36.5	36.0	35.7	35.4	35.2	35.0
W_{air} relative to that for an 8 m.e.v. α particle.	1.103	1.066	1.043	1.028	1.020	1.011	1.006	(1.000)

Table V shows the manner in which W probably varies with energy for α particles and protons, as computed by Gray (*23*) from a survey of experimental data. The mean initial energy of the protons may be taken to be half that of the neutrons under consideration.

In attempting to match the atomic composition of the chamber wall with that of the tissue being irradiated the important element to consider is the hydrogen, since this usually gives rise to rather over 90% of total energy absorption. It will generally be impossible to obtain an exact match but, when interpolating between two substances, one more rich and one less rich in hydrogen than tissue, advantage may be taken of the fact that the energy absorption increases linearly with hydrogen content. Aebersold's observations showed that neutron energy absorption in amber is very close to that in tissue. The figures given in Table VI are taken from Aebersold and Anslow's paper (*91*). It is clear that the wall thickness of the ionization chamber must be at least equal to the range of the most energetic recoil protons, and similarly that in the biological irradiations the

specimen must be surrounded on all sides by material having the same hydrogen content as itself.

TABLE VI

Relative Dose in Equivalent Roentgens of Various Materials Exposed to the Same Beam of Neutrons Generated by 8 m.e.v. Deuterons Bombarding Beryllium (91)

Material	Composition	Energy absorbed per g. material exposed to 8 m.e.v. neutrons[a]
Tissue (Aebersold and Anslow)........	$C_{0.5}H_5O_{3.8}N_{0.14}$	1.00
Tissue (Gray and Read)................	$C_{0.8}H_8O_4$	0.94
Paraffin...........................	CH_2	1.34
Amber.............................	$C_5H_8O_{0.5}$	1.03
Bakelite...........................	$C_7H_8O_2$	0.70
Celluloid..........................	$C_{5.3}H_8O_{3.5}N_{0.7}$	0.64
Air...............................	$O_{0.4}N_{1.6}$	0.19_5
Acrion (Zimmer)...................	$C_{10.9}H_8O_{2.7}$	0.53

[a] Arbitrary units.

Since the neutron output of a source is usually quoted in the physical literature in terms either of the total number of neutrons emitted in all directions or of the neutron flux at a given distance from the source, it may be convenient to note the following relation between the energy absorbed per unit mass of a medium, E_m, and the number, N, of neutrons incident per unit area, namely:

$$E_m = NE \ \Sigma \ n_i \ \sigma_i \ \frac{2A_i}{(1 + A_i)^2} \tag{20}$$

where E is the energy of individual neutrons, n_i is the number of atoms of type i per unit mass of the medium, σ_i is the atomic scattering cross section of atoms of type i for neutrons of energy E, and A_i is the atomic weight of atoms of type i. The sign Σ indicates that the contributions of all types of atom are to be added together. The formula implicitly assumes that the scattering of neutrons by all types of atom is isotropic, which is not strictly true. It is, however, true of hydrogen, which contributes about 92.5% of the energy absorption in tissue. Indeed a good approximation to the energy absorption for neutrons of any given energy may be obtained by multi-

plying the hydrogen term in the energy absorption by $100/92.5 = 1.08$. Thus, very simply, we have the approximate relation:

$$E_m = 1.08 N E n_{\mathrm{H}} \sigma_{\mathrm{H}} \times 0.5 = 0.54 N E n_{\mathrm{H}} \sigma_{\mathrm{H}} \qquad (21)$$

If we wish to convert to dose in equivalent roentgens, the expression (21) must be divided by 84 if energies are expressed in ergs or 5.23×10^{13} if measured in electron volts. To convert energy units, divide by 93 or 5.81×10^{13}, respectively.

Table VII gives the dose in equivalent roentgens due to a flux of N neutrons per square centimeter of the stated energy, calculated from published experimental values of σ_{H} for neutrons (23). The dose per unit neutron flux evidently varies relatively slowly with neutron energy.

TABLE VII

Approximate Dose Received by Tissue Exposed to Neutron Flux of 10^9 Neutrons per Square Centimeter as a Function of Neutron Energy

Neutron energy, m.e.v.	0.9	2.44	2.88	4	25
Dose, r.e.p.	2.03	3.39	4.15	4.23	5.95

Slow and Thermal Neutrons. So little is known about the ionization and excitation caused by neutrons having less than about 0.2 m.e.v. energy, which we have classified as slow neutrons, that their measurement must be regarded as a research in itself. It is fairly clear, however, from a consideration of alternative modes of absorption of thermal neutrons in tissue that when neutrons of energy less than about 0.05 m.e.v. fall on tissue the energy associated with the disintegration of the nitrogen content of the tissue, which is produced by some 8% of these neutrons, will contribute at least as much energy as is contributed in all other forms in the course of the gradual slowing down of the neutrons. Thus, in the case of very slow neutrons and thermal neutrons, we may concentrate attention on the disintegration of the nitrogen in the tissue element under consideration and the irradiation of this element by the 2.2 m.e.v. γ rays which come from the absorption of thermal neutrons elsewhere.

The relation between energy absorption per unit mass of tissue, E_m, and thermal neutron flux, N, is given by an equation analogous to equation (21), namely.

$$E_m = N n_{\mathrm{N}} \sigma_{\mathrm{N}} E_{\mathrm{N}} \qquad (22)$$

where n_N is the number of nitrogen atoms per gram of tissue, σ_N is the atomic cross section for disintegration, and E_N is the energy release per disintegration. Assuming 2.5% by weight of nitrogen in tissue, $\sigma_N = 1.7 \times 10^{-24}$ cm.² and $E_N = 0.62$ m.e.v. (94), the dose is related to the neutron flux by:

$$\text{dose} = 2.18 \times 10^{-11} \, N \text{ r.e.p.} \qquad (23)$$

The most direct method of measuring the dose due to slow neutrons is to measure the ionization produced in air in a chamber large compared with the range of the disintegration protons—*i.e.*, large compared with 1 cm. The ionization observed will be a measure of the energy loss in a medium in which the percentage of nitrogen by weight is the same as in air, namely, 75.5%. The dose received by tissue is therefore related to the ionization per unit volume of air at 0°C. and 760 mm. of mercury by:

$$\text{dose} = 1.11 J_v \times (p/0.75_5) \text{ r.e.p.} \qquad (24)$$
$$= 1.0 J_v \times (p/0.75_5) \text{ energy units}$$

where p is the proportion by weight of nitrogen in tissue. The factor 1.11 is the ratio of the energy expended per ion pair formed in air by disintegration protons relative to fast electrons. When the dose is expressed in energy units, this factor is exactly balanced by the ratio of the electron densities of tissue and air.

Mixed Radiations. The measurement of the dose in equivalent roentgens received by tissue exposed to mixed γ-ray, fast neutron, and slow neutron beams presents no new problem in principle. It requires only that a chamber be constructed in which the elements vital to the correct recording of neutron dose, namely, hydrogen and nitrogen, should be present in the chamber wall and in the gas in correct proportions. If this is accomplished, the chamber (with appropriate allowance for W, the average energy expended in the production of a pair of ions in the gas mixture) will correctly record the dose due to all three types of radiation together in equivalent roentgens, except for the energy dissipated by epithermal neutrons, which has no proportional ionization associated with it. This omission is not likely to be serious in practice.

It unfortunately happens, however, that, in the practical problem of monitoring research establishments to ascertain that the permissible dose levels are not exceeded, the total effect as regards injury to health of the ionization due to fast and slow neutrons is weighted

ten times in relation to γ-ray ionization. No way has so far been
devised to accomplish this weighting fully except by analysis of the
observed ionization into the components due to γ radiation on the one
hand and fast and slow neutron ionization on the other.   With care
this may usually be accomplished by the use of at least two different
ionization chambers, the one of which has true tissue composition
and responds to all radiation proportionately, and the other having
no hydrogen or nitrogen content and therefore heavily weighting γ
ray as compared with the other forms of energy absorption.   Inas-
much as hydrogen, which is present only to the extent of 10% (Table
VI) contributes over 90% of the neutron energy absorption in tissue,
it is clear that the neutron energy absorbed per gram will be ten times
greater in hydrogen than in tissue.   Gamma-ray energy absorption
per gram is twice as great in hydrogen as in any other element.   A
large ionization chamber containing hydrogen gas at high pressure
therefore automatically weights neutron ionization relative to γ-ray
ionization in the ratio of about five to one.   Boron might be added
to make an appropriate slow neutron contribution.   The experimental
separation of γ-ray and fast neutron effects was studied by Lea (*95*)
and by Aebersold and Anslow (*91*).

**Beta Radiation and Alpha Radiation from Radioactive Material
Uniformly Distributed in Tissue.**   If the concentration of radio-
active material in tissue is known and also the mean energy of the
β particles emitted at each disintegration, it is obviously easy to cal-
culate the average β-ray energy expended per unit mass of tissue in
which a radioactive element is uniformly incorporated, and if this
energy is divided by 84 ergs per gram we obtain the dose in equiva-
lent roentgens.   The total dose resulting from complete decay of the
radioactive material has been given in a convenient form as follows
(*55*):

$$D_\beta = 88 \; \bar{E}_\beta \; TC \text{ r.e.p.} \tag{25}$$

$$= 79 \; \bar{E}_\beta \; TC \text{ energy units}$$

where $\bar{E}_\beta$ is the mean energy in m.e.v. of the β rays emitted by this
source, T is the half-life in days, and C is the concentration of the iso-
tope in microcuries per gram

Correspondingly, the initial dose rate is given by:

$$\text{dose rate} = 2.54 \; \bar{E}_\beta C \text{ r.e.p./hour} \tag{26}$$

$$= 2.29 \; \bar{E}_\beta C \text{ energy units/hour}$$

The same formula may be used for a source that emits a single α particle at each disintegration if the full energy of the α particle is substituted for $\bar{E}_\beta$. $\bar{E}_\beta$ has been evaluated (55) for a considerable number of radioelements so that if C, the concentration of element in microcuries per gram is known, the dose rate in the tissues may be evaluated.

Unfortunately, there is still much disagreement among different laboratories in the estimation of the strength of a given source, and the method of estimating source strength by $\beta-\gamma$ coincidences is not applicable to one of the commonest elements, P^{32}, because this isotope emits no γ rays. It is therefore worth noting that if a radioelement is uniformly incorporated in the material of the wall of an ionization chamber the ionization recorded is related to the energy absorbed per gram of wall material in exactly the same way as X-ray ionization and neutron ionization are related to the absorption of primary radiation energy in the chamber wall material. As is the case with X rays it is almost always possible to choose the size of chamber and the gas pressure so that air may be used, and the necessity for special gas fillings avoided. Failla (49, p. 637) has investigated the possibility of using the "extrapolation ionization chamber" for the measurement of the concentration of radioactive material in solutions, and Gray (101) has measured the rate of emission of β-ray energy by a number of isotopes by cavity-type ionization chambers. The ionization method may well prove to be one of the most satisfactory methods of estimating dose in those cases in which uncertainty exists concerning the form of the β-ray spectrum, for estimating concentration of radioactive material, and even for determination of the absolute strength of sources, in millicuries, by the application of equation (26) in those cases in which the form of the β-ray spectrum has been well established by β-ray spectroscopy.

5. Ionization in Tissue

In the earlier sections of this chapter we frequently spoke of the number of ions formed in tissue by a given dose of radiation and of the number of ions per micron of track. This was done because it is very helpful to our understanding of the biological effects of radiation to visualize the approximate position of the ions, and because there are experimental grounds for believing that at least in some cases (cf. the discussion of the production of mutations to phage resistance

in *Escherichia coli*, Sect. A6) the ions are the effective agent.   On a number of occasions it has been emphasized that excitation should not be overlooked.

From the considerations advanced in this section, it is clear that the quantity we are able to estimate by the appropriate use of ionization chambers is not the amount of ionization produced in the tissue but the total energy in all forms – ionization, excitation, chemical change, and thermal agitation—transferred by the ionizing particles to the tissue.   So far no one has even demonstrated that the ions H_2O^+ and H_2O^- are in fact formed by the irradiation of liquid water, though on theoretical grounds there can be no doubt as to their formation.   These ions are demonstrable in the gaseous phase and the minimum energy of an electron capable of ionizing water vapor is known to be 13 e.v.   The presence of water vapor in ionization chambers is usually studiously avoided for the sake of maintaining the essential high grade insulation between collecting electrode and earth so that until recently no measurements had been made of W_{H_2O}, the mean energy expended in the production of a pair of ions in water vapor.   Appleyard (*102*) finds $W_{water vapor}/W_{air} = 0.88 \pm 0.02$ for α particles of 4.5 m.e.v. initial energy.   Since the value in other gases is usually about double the ionization potential, the value of W_{H_2O} for electrons would be expected to be about 26 e.v.

It would seem that, whereas dose may appropriately be specified in some quantity such as roentgens, equivalent roentgens, or energy units directly related to ergs of energy absorbed per gram of tissue, it is useful to recognize that there are probably formed about two ions in each cubic micron of tissue exposed to 1 equivalent roentgen of any kind of ionizing radiation.

References

1. Lea, D. E., *Actions of Radiations on Living Cells*.   Cambridge Univ. Press, Cambridge, 1946.

2. Kotval, J. P., and L. H. Gray, *J. Genetics*, **48**, 135 (1947).

3. Figge, F. H. J., *Science*, **105**, 323 (1947).

4. The Plutonium Project, *Radiology*, **49**, 269 (1947).

5. Henshaw, P., *J. Nat. Cancer Inst.*, **1**, 789 (1941).

6. Raper, J. R., *Radiology*, **49**, 314 (1947).

7. Henshaw, P., E. F. Riley, and G. E. Stapleton, *Radiology*, **49**, 349 (1947).

8. Lorenz, E., W. E. Heston, A. B. Eschenbrenner, and M. K. Deringer, *Radiology*, **49**, 274 (1947).

9. Lisco, H., M. P. Finkel, and A. M. Brues, *Radiology*, **49**, 361 (1947).
10. Muller, H. J., *Z. indukt. Abstamm.- u. VererbLehre*, Supplement No. 1, 234 (1928).
11. Zimmer, K. G., and N. W. Timofeeff-Ressovsky, *Strahlentherapie*, **63**, 528 (1938).
12. Giles, N. H., *Genetics*, **28**, 398 (1943).
13. Dempster, E. R., *Proc. Natl. Acad. Sci. U.S.*, **27**, 249 (1941).
14. Fano, U., *Genetics*, **28**, 74 (1943).
15. Demerec, M., B. P. Kaufmann, and E. Sutton, *Genetics*, **27**, 140 (1942).
16. Zimmer, K. G., and N. W. Timofeeff-Ressovsky, *Strahlentherapie*, **55**, 77 (1936).
17. Ward, F. D., *Genetics*, **20**, 230 (1935).
18. Gowen, J. W., *Cold Spring Harbor Symposia Quant. Biol.*, **9**, 187 (1941).
19. Demerec, M., and R. Latarjet, *Cold Spring Harbor Symposia Quant. Biol.*, **11**, 38 (1946).
20. *Cold Spring Harbor Symposia Quant. Biol.*, **11** (1946).
21. Lea, D. E., and M. H. Salaman, *Proc. Roy. Soc. London*, **B133**, 434 (1946).
22. Wilson, C. T. R., *Proc. Roy. Soc. London*, **A104**, 192 (1923).
23. Gray, L. H., *Proc. Cambridge Phil. Soc.*, **40**, 72 (1943).
24. Lea, D. E., R. B. Haines, and E. Bretscher, *J. Hyg.*, **41**, 1 (1941).
25. Spear, F. G., *Brit. J. Radiology*, **17**, 348 (1944).
26. Catcheside, D. G., and D. E. Lea, *J. Genetics*, **45**, 186, 1943.
27. Luria, S. E., and T. F. Anderson, *Proc. Natl. Acad. Sci. U. S.*, **28**, 127 (1942).
28. Lawrence, E. O., *Radiology*, **29**, 313 (1937).
29. Rutherford, E., J. Chadwick, and C. D. Ellis. *Radiations from Radioactive Substances*, Cambridge Univ. Press, London, 1930.
30. Kara-Michailova, E., and D. E. Lea, *Proc. Cambridge Phil. Soc.*, **36**, 101 (1940).
31. Lea, D. E., and M. H. Salaman, *Brit. J. Exptl. Path.*, **23**, 27 (1942).
32. Kotval, J. P., Thesis, Cambridge, 1944.
33. Dale, W. M., W. J. Meredith, and L. H. Gray, *Proc. Roy. Soc. London*, **242**, 33 (1949).
34. Bethe, H., in *Handbuch der Physik*. Vol. 24, Part 1, Springer, Berlin, 1933, p. 273.
35. Hornbeck, G., and I. Howell, *Proc. Am. Phil. Soc.*, **84**, 33 (1941).
36. Shearin, P. E., and T. E. Pardue, *Proc. Am. Phil. Soc.*, **85**, 243 (1942).
37. Alper, T., *Z. Physik*, **76**, 172 (1932).
38. Klemperer, O., *Z. Physik*, **45**, 225 (1927).
39. Nurnberger, C. E., *Proc. Natl. Acad. Sci. U. S.*, **23**, 189 (1937).
40. Stenstrom, W., and A. Lohman, *J. Biol. Chem.*, **79**, 673 (1928).
41. Duane, W., and O. Scheuer, *Radium*, **10**, 33 (1913).
42. Frilley, M., *Brit. J. Radiology*, Supplement No. 1, 50 (1947).

43. Fricke, H., and E. R. Brownscombe, *Phys. Rev.*, **44**, 240 (1933).

44. Weiss, J., *Nature*, **153**, 748 (1944).

45. Marton, L., and P. H. Abelson, *Science*, **106**, 69 (1947).

46. Cooper, F. S., C. E. Buchwald, C. P. Haskins, and R. D. Evans, *Rev. Sci. Instruments*, **10**, 73 (1939).

47. Morningstar, O., R. D. Evans, and C. P. Haskins, *Rev. Sci. Instruments*, **12**, 358 (1941).

48. Gerthsen, C., *Ann. Physik*, **5**, 657 (1930).

49. Medical Physics. O. Glasser ed., Year Book Publishers, Chicago, 1944.

50. Trump, J. G., R. J. Van de Graaff, and R. W. Cloud, *Am. J. Roentgenol.*, **43**, 728 (1940).

51. Robbins, L. R., J. C. Aub, O. Cope, D. Cogan, J. L. Langohr, R. W. Cloud, and O. E. Merrill, *Radiology*, **46**, 1 (1946).

52. Skaggs, L. S., G. M. Almy, D. W. Kerst, and L. H. Lanzl, *Phys. Rev.*, **70**, 95 (1946).

53. Kingdon, K. H., P. A. Zahl, C. P. Haskins, and H. E. Tanis, *Radiology*, **31**, 52 (1938).

54. Brasch, A., and W. Huber, *Science*, **105**, 112 (1947).

55. Marinelli, L. D., R. F. Brinckerhoff, and G. J. Hine, *Rev. Mod. Phys.*, **19**, 25 (1947).

56. Neary, G. J., *Brit. J. Radiology*, **19**, 357 (1946).

57. McMillan, E. M., *Phys. Rev.*, **68**, 143 (1945).

58. Veksler, V., *J. Phys. (U. S. S. R.)*, **9**, 153 (1945).

59. Wilson, R. R., *Radiology*, **47**, 487 (1946).

60. Chadwick, J., *Proc. Roy. Soc. London*, **136**, 692 (1932).

61. Bothe, W., and H. Becker, *Z. Physik*, **66**, 289 (1930).

62. Gray, L. H., J. Read, and J. G. Wyatt, *Brit. J. Radiology*, **13**, 82 (1940).

63. Gray, L. H., and J. Read, *Brit. J. Radiology*, **13**, 248 (1940).

64. Aebersold, P. C., and J. H. Lawrence, *Ann. Rev. Physiol.*, **4**, 25 (1942).

65. Aebersold, P. C., *Phys. Rev.*, **56**, 714 (1939).

66. Stone, R. S., and J. C. Larkin, *Radiology*, **39**, 608 (1942).

67. Zirkle, R. E., *Radiology*, **49**, 271 (1947).

68. Gray, L. H., and J. Read, *Brit. J. Radiology*, **21**, 5 (1948).

69. Bonner, T. W., *Phys. Rev.*, **52**, 685 (1937).

70. Krueger, A. P., *J. Gen. Physiol.*, **13**, 557 (1930).

71. Zahl, P. A., F. S. Cooper, J. R. Dunning, *Proc. Natl. Acad. Sci. U. S.*, **26**, 589 (1940).

72. Livingston, M. S., and H. A. Bethe, *Rev. Mod. Phys.*, **9**, 245 (1937).

73. Mitchell, J. S., *Brit. J. Radiology*, **20**, 79 (1947).

74. Amaldi, E., D. Bocciarelli, F. Rasetti, and G. C. Trabacchi, *Phys. Rev.*, **56**, 881 (1939).

75. Lea, D. E., R. B. Haines, and C. A. Coulson, *Proc. Roy. Soc. London*, **B120**, 63 (1936).

76. Zirkle, R. E., *J. Cellular Comp. Physiol.*, **16**, 221 (1940).

77. Gray, L. H., and J. Read, *Brit. J. Radiology*, **15**, 320 (1942).

78. Spear, F. G., K. Tansley, and L. H. Gray, *Brit. J. Radiology*, **21**, 567 (1948).

79. Bonet-Maury, P., *Brit. J. Radiology*, Suppl. 1, 21 (1947).

80. Colwell, H. A., and S. Russ, *Radium, X-Rays and the Living Cell*, 2nd ed. Bell, London, 1924.

81. Gray, L. H., and J. Read, *Nature*, **152**, 53 (1943).

82. Evans, R. D., *Nucleonics*, **1**, 32 (1947).

83. Gray, L. H., *Proc. Roy. Soc. London*, **A159**, 263 (1937).

84. Rutherford, E., *Radioactive Substances and Their Radiations*. Cambridge, Univ. Press, London, 1913.

85. Wilson, C. W., *Radium Therapy*. Chapman & Hall, London, 1945.

86. Farmer, F. T., *Brit. J. Radiology*, **19**, 27 (1946).

87. Lea, D. E., *J. Sci. Instruments*, **14**, 89 (1937).

88. Sievert, R., *Acta Radiol.*, Suppl. 14 (1932).

89. Smereker, H., and K. Juris, *Strahlentherapie*, **52**, 327 (1935).

90. Bragg, W. H., *Studies in Radioactivity*. Macmillan, London, 1912.

91. Aebersold, P. C., and G. A. Anslow, *Phys. Rev.*, **69**, 1 (1946).

92. Zimmer, K. G., *Strahlentherapie*, **63**, 517 (1938).

93. Gray, L. H., and J. Read, *Nature*, **144**, 439 (1939).

94. Baldinger, E., and P. Huber, *Nature*, **143**, 894 (1939).

95. Lea, D. E., *Proc. Roy. Soc. London*, **A150**, 637 (1935).

96. Reddemann, H., *Z. Physik*, **110**, 373 (1938).

97. Amaldi, E., L. R. Hafstad, and M. A. Tuve, *Phys. Rev.*, **51**, 896 (1937).

98. Preer, J. R., *Am. Naturalist*, **82**, 35 (1948).

99. Lawrence, J. H., W. F. Loomis, C. H. Tobias, and F. H. Turpen, *J. Physiol.*, **105**, 197 (1946).

100. Spiers, F. W., and G. W. Blomfield, *Brit. J. Radiology*, **19**, 349 (1946).

101. Gray, L. H., *Brit. J. Radiology*, in press.

102. Appleyard, R. K., private letter to be communicated to *Nature*.

103. Allen, A. O., *J. Phys. & Colloid Chem.*, **52**, 479 (1948).

104. Lea, D. E., *Brit. J. Radiology*, Suppl. No. 1, 59 (1947).

105. Appleyard, R. K., *Nature*, **163**, 526 (2949).

STABLE ISOTOPES AS TRACERS

Fred M. Uber, *Iowa State College*

A. FUNDAMENTAL BASIS OF STABLE ISOTOPE METHODS

1. Principles and Underlying Assumptions

Nearly three hundred different stable atoms have been discovered and classified on the basis of their mass. Since there are less than a hundred chemical elements in the periodic table, many elements must comprise atoms with two or more mass types. Atoms that differ in mass but not in their ordinary chemical behavior are called isotopes. Substances that possess only one stable isotope each are not generally useful for our present purpose. Some representatives of the twenty odd elements in this group are: sodium of mass 23 (usually written as Na^{23}), Al^{27}, P^{31}, and I^{127}. Biologically important elements that consist of two stable isotopes each are hydrogen, carbon, nitrogen, and chlorine. Among other elements useful in biological investigations, mention must be made of oxygen, which has three stable isotopes, and of sulfur, which has four.

When one of the two or more stable isotopes of a chemical element occurs naturally with a very low concentration, the possibility exists of employing this rare isotope as an atomic marker or label. It is necessary, however, that its natural concentration first be in-

TABLE I

Natural Concentration of the Stable Isotopes Most Commonly Employed in Biological Studies

Element	Mass	Atom per cent concn.	Mass	Atom per cent concn.	Mass	Atom per cent concn.	Mass	Atom per cent concn.
Hydrogen	1	99.98	2	0.02	..	..	..	..
Carbon	12	98.9	13	1.1	..	..	..	..
Nitrogen	14	99.62	15	0.38	..	..	..	..
Oxygen	16	99.76	17	0.04	18	0.20	..	..
Sulfur	32	95.1	33	0.74	34	4.2	36	0.016

creased artificially. Since practical processes for segregating isotopic components have been developed in recent years (1,3,6), this labeling possibility has now been realized for the elements listed in Table I. The basis then for the use of stable isotopes as atomic labels with which to observe the behavior of molecules consists of synthesizing compounds in which the comparatively rare isotope of an element occurs in a greater than normal concentration.

As an example, let us suppose that we wish to investigate the source of the oxygen in the water obtained when ethyl alcohol reacts with acetic acid to form ethyl acetate. For one possible approach, ethyl alcohol in which the oxygen isotope of mass 18 occurred in excess of its normal concentration could be synthesized and then allowed to react with ordinary acetic acid. An analysis of the water derived from this reaction would then be made to determine its O^{18} concentration. If its O^{18} content were found to be normal, one would conclude that the oxygen in the water came from the acetic acid and not from the ethyl alcohol. A confirmatory analysis could be made on the ethyl acetate. In this case, it has been established (15) that the oxygen in the water splits off from the acetic acid and the reaction is written as follows:

$$C_2H_5O^{18}H + CH_3COOH \rightleftharpoons CH_3COO^{18}C_2H_5 + H_2O$$
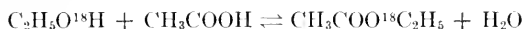
The biological use of the least concentrated isotope of an element as a tracer is based on the assumption that organisms are unable to distinguish between the two or more isotopes of a chemical element. The plausibility of this assumption is favored by the relatively late discovery of isotopes and by the extreme difficulties encountered in their concentration in the laboratory. It has now been established that organisms in the course of their normal metabolic processes are incapable of discriminating between the various isotopes of an element within rather wide limits (24). The only noteworthy exception occurs in the case of hydrogen. This is not surprising in view of the fact that the heavy hydrogen isotope is just twice the mass of the lighter one. Even with hydrogen no appreciable differential effect is shown biologically except in high concentrations. When it is realized that the stable isotopes used as labels are everywhere present in nature, only in somewhat smaller concentration, it is clear that nothing is being introduced that is essentially new or foreign to a normal living organism. A theoretical discussion of the validity of the use of tracers to follow chemical reactions recently appeared (24a).

2. Theory of Measurement

With the possible exception of the hydrogen isotopes, whose mass ratio of two is great enough to permit accurate determinations of concentration by means of density measurements, the instrument of choice for determining isotope concentrations is the mass spectrometer. When electrically charged particles enter a uniform magnetic

field perpendicular to their direction of motion, each particle experiences a deflecting force. This force is perpendicular both to the direction of the magnetic field and to the trajectory of the particle. As a consequence the charged particles are forced to move in circular paths (see Fig. 3). For a constant magnetic field and for particles of the same kinetic energy, the radius of curvature of any particle's path is determined by the ratio of its mass to its electrical charge. Hence, a mass spectrometer is an instrument for segregating ionized particles according to their mass-to-charge ratio.

When a sample of a gaseous compound is subjected to electronic bombardment in a mass spectrometer, usually several different ionized particles are produced (see Table II). All of these ions are accelerated by the same constant electrical potential and are then permitted to pass between magnetic poles. Each particle will be deflected by an amount characteristic of its mass-to-charge ratio. Consequently the various particles may be segregated into groups by having them pass through narrow exit slits to appropriate collectors (4, p. 398). Since each particle carries a charge, the electric current through each collector is a measure of the isotope concentration. Usually only one collector is employed and measurements on the various isotopes are obtained by altering the electrical accelerating potential. In this manner, ions having any desired mass-to-charge ratio may be deflected at will through a single exit slit into the collecting chamber.

3. When Should Stable Instead of Radioactive Isotopes Be Used?

Isotopes of several important tracer elements are available in both stable and radioactive forms. This is true of hydrogen, carbon, nitrogen, oxygen, and sulfur. Therefore, a decision must be made as to whether to employ a stable or a radioactive isotope for a particular problem. Some factors to be considered are: access to measuring equipment, relative cost and technical skill required, duration of the experiment, possibility of interference from radioactive radiations, relative difficulty of making measurements, and finally, but often most important of all, the dilution factor that may be encountered.

If apparatus is already in operation for making measurements on one type of isotope, for example a mass spectrometer used for N^{15}, a proposed investigation involving isotopic carbon might be made more

readily with the stable C^{13} than with either the radioactive C^{14} or C^{11}. On the other hand, if studies in progress with radioactive carbon were to be supplemented by some investigations with isotopic nitrogen, it would probably be necessary to secure a mass spectrometer so that N^{15} could be used rather than the radioactive N^{13}. The decisive factor here is the very short half-life of N^{13}, which is approximately ten minutes. It should also be borne in mind that shifting from one type of equipment to another involves learning a different technique and represents a loss of time unless it serves some useful purpose.

Some conception of the cost of equipment for measuring stable isotopes will be gained from a later section of this chapter; and for radioactive isotope equipment, from a separate chapter. A mass spectrometer will certainly cost more initially than equipment for radioactive isotope measurements. On the other hand a mass spectometer is useful for the several isotopes of greatest interest to the biologist without any additional modification or expense. Where only deuterium experiments are contemplated, the concentrations can be determined by density measurements on water so that the cost of the apparatus is almost negligible (1, pp. 51 and 61). For measurements on some of the perhaps most useful radioactive isotopes, namely, H^3, C^{14}, and S^{35}, special apparatus must be constructed, owing to the very weak radiations emitted. Requirements with respect to apparatus and technical skill for measuring the radioactive isotopes characterized by more penetrating radiation are lower than for the three elements mentioned above or for the operation of a mass spectrometer. Both types of equipment may require the services of an electronic expert when difficulties develop; and they will develop.

A distinct advantage that attaches to the stable isotopes is their stability. Where experiments must run over long intervals of time or where they involve time-consuming organic syntheses, it is comforting to know that the isotopes employed are not disintegrating. For example, the rapid decay rate of N^{13} and C^{11} practically eliminates their use for numerous experiments, although a few valuable studies have been made with the latter. C^{14} with a half-life of about 5000 years is essentially stable, while S^{35} (half-life 88 days) and H^3 (half-life 11 years) have decay rates suitable for most experiments.

Although the radiation of energy that attends the spontaneous disintegration of radioactive isotopes constitutes a necessary event from the standpoint of concentration measurements, it is conceivable

that the radiation itself may be a distinctly harmful factor biologically. This may well be true of genetic experiments or where one is dealing with cells that are actively dividing and therefore relatively sensitive to radiation. In many studies with tracers, the concentration of radioactive material need not be so high that great concern must be felt about the biological effects of the radiation emitted. Nevertheless, the matter must always be considered in relation to each individual experiment and a decision made on the merits of the several factors involved.

The technical skills involved in concentration measurements of both stable and radioactive isotopes have been discussed elsewhere. It will be noted that measurements on stable isotopes normally require a qualitative extraction of the biological material so that test samples of the various constituents can be obtained in a high degree of purity. While this is often true of radioactive isotopes also, there are some instances in which the path of transportation or the site of deposition of radioactive isotopes may be studied in a living organism without resorting to any chemical segregation of the material. Obviously this represents a time-saving simplification, which may well weight a decision in favor of a radioactive isotope for a given type of study.

Permissible dilution factors for the various stable isotopes have a limiting value in the neighborhood of 10,000. This is adequate, of course, for many investigations. However, the principal advantage of the radioactive over the stable isotopes lies in the fact that a much greater dilution of radioactive isotopes can be tolerated. Dilution factors as great as 1,000,000 or even 10,000,000 lie within the realm of possibility. Clearly this will extend the use of radioactive tracers to the investigation of compounds that occur in minute quantities only. Before arriving at a decision as to which type of isotope to employ in a given experiment, a careful estimate should be made of the dilution to be expected.

B. POTENTIAL RESEARCH APPLICATIONS

1. Diffusion and Transport Studies

The transport of nutrient materials, both organic and inorganic, in living organisms has been studied for many years by a great many ingenious methods. Most of these earlier techniques have been out-

moded by the wide range of application of isotopic tracers to transport problems. The factor primarily responsible for this ready acceptance of a new method lies in the incontrovertibility of the stable isotope evidence when properly obtained and evaluated. Nothing foreign to the biological system is being introduced, as has almost always been the case with labels of other descriptions. Numerous metabolic precursors and products that formerly were even impossible to label may now be followed satisfactorily in living systems.

It is to be expected that exploratory investigations by means of tracer isotopes will confirm many of the conclusions reached earlier from the use of other techniques. Contributions to date have been largely of this character. The superior potentialities inherent in the isotope tracer method should lead soon, however, to more extensive and detailed studies on the translocation of a multitude of substances, including important biochemical intermediates. A beginning has been made, especially with metallic radioactive ions (3,3b,17), but an immense field for research involving carbon compounds has barely been touched.

Transportation of the products of photosynthesis through plant tissues has been studied by several workers and an introduction to the literature on the subject may be found in an article by Rabideau and Burr (26). An initial study on sulfur transport in plants has also appeared (25). Significant future advances in our knowledge regarding the permeability of cells of all types seem inevitable now that steady state conditions can be studied so readily by virtue of isotopic labels. Here again, a beginning has been made (17, p. 118).

2. Mechanisms of Biochemical Reactions

Perhaps the most significant contributions resulting from the use of stable isotopes in biology have been concerned with the intermediary steps in metabolism. Hitherto proposed mechanisms for biochemical reactions were based primarily on circumstantial evidence obtained in devious ways. Direct evidence available from reactants labeled with stable isotopes has not only clarified and confirmed many of the older conceptions of intermediary metabolism but also has resulted in the introduction of a number of new ideas. The speed with which hypotheses in this field can be tested has been increased almost immeasurably and consequently progress can be expected to be rapid in the future.

Much valuable work has been done recently on the carbon dioxide metabolism that occurs almost universally in biological systems. Perhaps because this kind of reaction could not be readily studied except after the advent of carbon isotopes, it has enjoyed unusual popularity the past few years. Since the results have been discussed rather widely elsewhere, it seems sufficient here merely to refer to a recent review (28).

Nitrogen metabolism within the animal body has been investigated extensively by a number of workers, but especially in the long series of researches by Schoenheimer, Rittenberg, and associates. Reference must be made to their reviews (23,24) for an adequate account of this work, which covers not only a great many of the amino acids but also other nitrogen compounds.

3. Rates of Biological Processes

The use of isotope labels has facilitated rate measurements for numerous biological processes. A suggestive list of possibilities could include the following: the rate of absorption of nutrient constituents by various tissues, the excretion rate of waste products either across permeable membranes or from an organism as a whole, the rate of metabolic turnover for various elements and compounds, the rate of deposition or accumulation of specific substances in definite locations within a plant or animal, and the rate of circulation of body fluids.

Simple exchange reactions constitute a group of special interest with respect to the use of isotopes since their rates cannot be measured by any other method. Further, the presence of an exchange reaction that involves the isotopic element used as a label will often disqualify the use of a particular compound for a given investigation. This is discussed in greater detail in a later section.

Many biological studies are conducted in which the organisms are maintained in an equilibrium or steady state condition. Isotopes are particularly applicable in such cases since they permit interesting reaction rates to be measured quantitatively even in the absence of any demonstrable net change in chemical constitution of the system.

4. Quantitative Analysis

The quantitative analysis of a multicomponent solution by the stable isotope dilution technique furnishes an excellent illustration of

the value of isotopic tracers. The usefulness of this procedure has been amply demonstrated by Rittenberg and Foster (16), who made quantitative analyses for a number of the amino acids resulting from protein hydrolysis, and by others (see 20, p. 476). A central feature is the fact that the isotope dilution method does not require the quantitative extraction of the components to be analyzed. It is only necessary to provide a sample containing one or perhaps a few milligrams of a test substance in a pure state. This circumstance is especially important in dealing with amino acids, where it is relatively easy to obtain either quantitative fractions of impure compounds or nonquantitative fractions of pure compounds, but exceedingly difficult to obtain samples that are both quantitative and pure. However, before any substance can be determined quantitatively, a pure synthetic preparation of it containing an isotopic marker must be available.

As an illustration of the method, let us suppose that we have a solution of hydrolyzed protein and that we wish to determine its glycine content, X. A synthetic glycine sample of known weight, G, which contains an N^{15} concentration of C_0 atom per cent in excess of normal, is added to the hydrolyzate. The synthetic and the natural glycine form an inseparable mixture with a uniform distribution of the N^{15}-tagged molecules. Therefore, the isolation of a small but pure fraction of the total glycine will enable its diluted N^{15} concentration, C atom per cent excess, to be determined. Without resorting to a quantitative extraction, the weight, X, of the original glycine can be calculated from the following equation (17, p. 116).

$$X = G \left(\frac{C_0}{C} - 1\right) \frac{M_X}{M_G} \tag{1}$$

The correction factor M_X/M_G represents the ratio of the molecular weights of the two glycine samples when the actual percentage of N^{15} present is considered. Most work to date has neglected this correction; though small, it is not always negligible in a quantitative sense.

In practice the analysis of other amino acids is considerably more complicated owing to the occurrence in synthetic preparations of both optical isomers. But, even here, analysis is possible (2, p. 563) and the isotope dilution method must be regarded as a generally applicable procedure (see especially 22b).

Since concentrations C_0 and C can be determined with an accu-

racy of better than 1%, the over-all accuracy of the dilution method
depends primarily on the purity of the individual isolated fractions.
As already pointed out, these fractions need not be quantitative ex-
tractions and consequently greater latitude is permitted in seeking a
higher purity.   Where several components are to be analyzed from a
single solution, with all of them having been enriched by the same
isotopic element, all individual errors are reduced somewhat percent-
agewise owing to mutual cross contamination of their respective im-
purities (2, p. 565).

5. Gas Analysis

To date, the use of the mass spectrometer for gas analysis has
been confined almost exclusively to the petroleum industry (7,18).
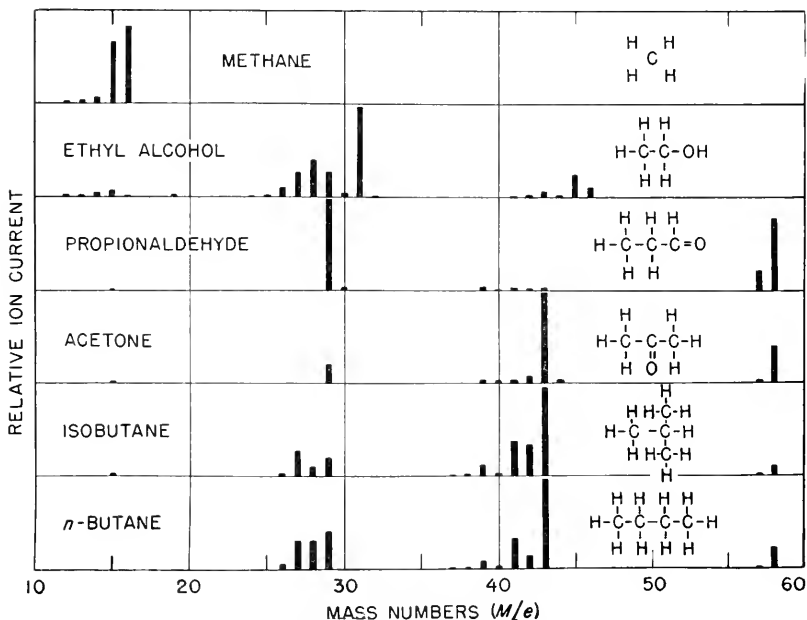
Fig. 1.   Typical mass spectra of a few organic molecules.

When mass spectrometers are more commonly available, it would seem
that they might be used extensively for gas analysis in connection
with biological problems.   Their ability to analyze 0.1 cc. (NTP)
gas samples quantitatively for a number of hydrocarbons should cer-

tainly find biological application. Typical mass spectra for a few common carbon compounds are shown in Figure 1. Note the characteristic variation in the relative concentrations of different ionized fragments, especially for the isomers of butane. Although automatic mass spectrometers for this purpose are admittedly complicated and expensive, there is a possibility in some cases that access may be had to one of the many instruments of this type already in operation at industrial research laboratories (see list in 1, p. 48).

It should be pointed out that quantitative analyses of small complex gas samples usually must rest on a previous calibration for each of the various components of the sample. This necessary complication, often readily surmountable with low molecular petroleum samples, appears formidable for biological compounds of high molecular weight. To be analyzed in the usual mass spectrometer (see Fig. 3), all liquid compounds should be appreciably volatile at room temperature. In specialized types of gas analysis, where one may be interested in determining the concentration of only one or two constituents of a complex sample, the problem is much more amenable to analysis. For example, an impurity of even one part in a million can be detected in commercial cylinders of compressed oxygen. Another related application of this technique is the location of leaks in metal vacuum systems. For this latter purpose it is customary to use a very much simplified form of mass spectrometer (8,12) and to probe suspected areas of the apparatus for leaks with helium. In this way leaks may be localized quickly without undesirable contamination and vexatious delays.

C. DEMANDS ON EQUIPMENT AND TECHNICAL SKILL

1. Availability of Concentrated Isotopes

Earlier workers in this field were forced to concentrate their own isotopes or to beg supplies from others who did. Several laboratory methods for the concentration of isotopes have been carefully worked out and are described in great detail in the literature (see 1,3,6). Should commercial supplies not be available, it is quite feasible to set up the necessary equipment and to concentrate a particular isotope in the laboratory. Usually this is a very involved undertaking. It should not be attempted with a viewpoint to obtaining isotopes at a

financial saving, particularly if there is any further reduction in the commercial prices. If one were to do so, it is not at all improbable that a year or more of an investigator's research time would be thus diverted. So far, the price of isotopic compounds, based on the excess weight of the artificially concentrated isotope present, has been quoted in terms of hundreds of dollars per gram. Since experiments can often be made with quite small quantities of starting material, the net cost per experiment is not really prohibitive.

Governmental distribution of deuterium through the United States Atomic Energy Commission (via *32*) has recently been announced. It may also be obtained commercially for research purposes (*35*) in the form of heavy water (D_2O). N^{15} has been made available commercially (*33*) in concentrations as high as 60 atom per cent excess. The N^{15} occurs in the ammonium radical of ammonium nitrate or in potassium phthalimide. Since relatively lower prices prevail for the lower concentrations, it is more economical to order the desired concentration initially than to obtain it later by direct dilution. C^{13} is also available (*33*) in concentrations as high as 60 atom per cent excess. No commercial source of stable sulfur isotopes is known to the writer at present but separation methods have been developed (*1*, p. 9). Limited quantities of isotopic oxygen can be purchased through the Atomic Energy Commission (*36*).

The number of synthetic organic compounds utilized in the past has been limited by the nature of the starting materials available and by the necessity of working on a micro scale. When stable isotopes become more plentiful and cheaper, it will be possible to purchase a wide variety of organic compounds that already contain one or more isotopic labels in specified positions. One manufacturer has already announced an intention to supply such compounds.

2. Synthesis of Isotopic Compounds

For the most part the ordinary synthetic methods of organic chemistry are utilized for the preparation of compounds containing isotope tracers. There is often a change in emphasis, however, such as the calculation of the yield on the basis of the isotopic starting compound rather than on some less expensive but more customary raw material. In synthesizing amino acids, for example, the crucial yield is the one based on ammonia since it contains N^{15}, rather than the yield related to the initial carbon chain. For this reason some N^{15} amino acids have been prepared more economically by the

phthalimide procedure of Gabriel or by the catalytic reduction of
α-keto acids according to Knoop and Oesterlin rather than by more
usual methods (24, p. 224).

For deuterium compounds, a very useful synthetic method con-
sists of the hydrogenation of double bonds. Examples of this process
are the synthesis of propionic acid from acrylic acid, stearic acid
from linoleic acid, butyric acid from crotonic acid, etc. (24). These
and other useful syntheses for numerous compounds containing
deuterium are included in a recent bibliography (17, p. 142). Hot
concentrated deuteriosulfuric acid is valuable in introducing deute-
rium on the α-carbon atom of fatty acids, and it may also be used for
amino acids. For some of the fatty acids it is necessary to use a
platinum catalyst in the presence of heavy water. Examples are
palmitic, lauric, capric, caprylic, and isocaproic acids. But with a
platinum catalyst, the deuterium atoms are equally distributed
among all the carbon atoms.

A widely applicable preparative method consists of feeding plants
or animals simple labeled substances and allowing them to synthesize
the desired compounds. Apart from the fact that the yields obtained
are often quite low, there is also a further dilution of the isotope with
respect to any given location in a molecule, since the atomic label
may appear in many positions. This method defeats its own purpose
in many cases since the label may not be sufficiently specific. Where
animals are permitted to drink heavy water, for instance, practically
all of the chemical constituents of the animal body will contain deu-
terium. But the deuterium may appear in several positions in each
of various molecules. A more direct biological synthesis can some-
times be employed by letting an organism convert one isotopic com-
pound into a second substance that would otherwise be much more
difficult to synthesize in the chemical laboratory. One example of
this is the direct conversion of stearic acid containing deuterium into
deuterio-oleic acid within the body of a mouse (24, p. 224).

Information about the synthesis of compounds containing carbon
isotopes can be obtained from contemporary research articles on
either C^{11}, C^{13}, or C^{14}. Two rather recent summaries have been fur-
nished (17, pp. 153–160; 22a) but only frequent contact with the cur-
rent periodical literature can keep one up-to-date (see 3, p. 161). A
similar remark applies to research with sulfur, in which much of the
work has been done with the radioactive S^{35}. Some syntheses with
isotopic sulfur have also been discussed by Kamen (17, pp. 205–212).

3. Apparatus for Detecting Stable Isotopes

Measurements on all of the commonly employed stable isotopes, with the possible exception of deuterium, are best made with the mass spectrometer. Owing to the fact that the mass of deuterium is just double that of ordinary hydrogen, pure samples of heavy water (D_2O) possess a density about 10% greater than ordinary water. Inasmuch as density measurements on liquids can be made with an accuracy of a few parts per million, it is feasible to deduce deuterium concentrations from the density of water samples. The advantages lie in the direction of simplicity and economy of apparatus. The latter may consist either of equipment for falling-drop determinations or of a gradient tube. The use of the equipment has been described adequately elsewhere (1, pp. 51–65).

There are several disadvantages to the method of density determinations on water as a means of determining deuterium concentration. Perhaps the greatest difficulty lies in achieving adequate purification of a water sample, since clearly any trace of impurity might influence the results very markedly. Purification then, and not density determination, is the main problem. A disturbing factor is the possible occurrence of any one of the three oxygen isotopes in a given water molecule. Since the heavy oxygen isotopes are usually concentrated in varying degree in commercial heavy water samples, this problem is one of some importance. A third disadvantage to the density determination method resides in the relatively large test sample required. Usually measurements are made on drops of 100 mg. This quantity exceeds that required with the mass spectrometer by roughly two orders of magnitude. However, where a research program involves only the deuterium isotope, and where samples of sufficient size can be made available for analysis, selection of the density method would seem to be logical.

The type of mass spectrometer employed in most biological studies to date was developed by Nier (3, p. 80; 4) at the University of Minnesota and is illustrated schematically in Figure 2.

The gas to be analyzed is contained in the 50 or 100 ml. flask, X, which is fitted with a stopcock and an interchangeable ground glass joint. (For an alternative device, see 11.) After evacuation of the capillary connecting tubes and the reservoir, Y, the test sample is transferred into Y and then compressed to the desired pressure (a few centimeters of mercury) as indicated by the attached manometer. Air is pumped into V by the rubber atomizer bulb, Z, if the height of the mercury in Y needs to be raised above

the barometric level. The evacuated storage flask W has a volume of 10 to 12 liters and is convenient for lowering the mercury level in Y. The gas
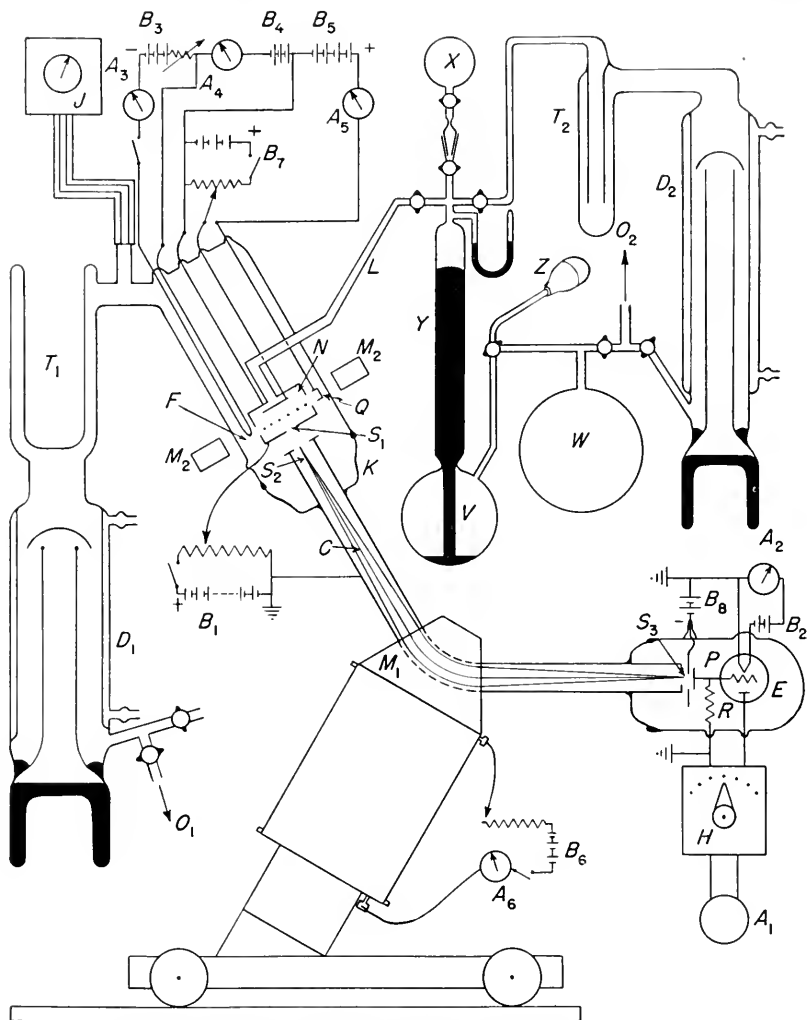
Fig. 2. Schematic diagram of a Nier type mass spectrometer, including electrical circuits and auxiliary equipment.

from Y is admitted slowly and continuously into the mass spectrometer ionization box, N, through the very fine leak, L (13).

Current to filament F is supplied by storage battery B_3 and indicated by

ammeter A_3. The electrons emitted by the filament traverse box N, as indicated by the dotted line, owing to an accelerating potential of 90 v. supplied by battery B_4. Most of the electrons are trapped on plate Q, which is about 90 v. more positive than N. The electron current to Q, perhaps 20 μamp., is read on meter A_5 and the total electron current to both Q and N, perhaps 40 μamp., is read on A_4. A magnetic field in the direction of the electronic flow, provided by the magnet M_2, either permanent or electric, confines the electrons to a rather well defined beam.

The ionized particles produced by the electronic bombardment acquire a small velocity in the direction of slit S_1 from the accelerating voltage supplied by battery B_7. But the principal accelerating potential difference of nearly 1000 v. is furnished by battery B_1 or a suitable rectifier (14), and occurs between exit slit S_1 in box N and entrance slit S_2 of C. In traversing the length of the long copper tube, C, the ions are deflected 60° by the sectorial magnetic field of the electromagnet, M_1, and brought to a focus at slit S_3. Upon emerging from C, the ions are collected on plate P. To prevent scattered electrons from the edges of slit S_3 from reaching P, a small negative potential from battery B_8 is applied to an intervening slit plate.

Because of the charged particles collected by P, and the resultant current through R, the grid of electrometer tube E experiences a change in potential. The resistor, R, has a resistance in the neighborhood of 10^{10} ohms. Usually two or more such resistors having different values are arranged so that either may be used at will in order to change the sensitivity. The calibrated variable amplifier, H, may include a circuit as described by Penick (9) or perhaps an inverse feedback type of amplifier (4,6,13). A wide range of sensitivities is desirable since relative isotope concentrations may vary by a factor of several hundred. In addition an automatic switching arrangement (5) should be provided to facilitate alternating the observations rapidly between the two or three mass peaks that characterize a particular test gas.

The operating gas pressure is less than 10^{-5} mm. of mercury, and should be read by an indicating gage, J (10,12), or by a McLeod gage located between trap T_1 and diffusion pump D_1. The trap is kept cold with either a solid carbon dioxide mixture or liquid air, depending on the material being tested. Mercury vapor diffusion pumps seem to be most satisfactory. Small rotary oil vacuum pumps, O_1 and O_2, precede the diffusion pumps.

Magnet M_1 may be wound for operation from either 12 v. storage batteries or a rectifier (4), or it may be replaced by a permanent magnet (6,8). Electronically regulated power supplies may be substituted for batteries in any of the indicated circuits. To facilitate baking out the spectrometer tube, the main magnet, M_1, is mounted on a carriage. It may be rolled to the left in the figure, thus enabling heating elements to be placed in position around metal tube C. Kovar collars, K, facilitate copper to Pyrex Vacuum seals.

Most mass spectrometers built have had deflection angles of

either 60, 90, or 180°. An instrument of the 90° type, but otherwise closely similar to the 60° design of Figure 2, was once built commercially and has been described by Hipple (2, p. 551). The popular 180° design originated by Dempster has been adopted for the automatically recording commercial mass spectrometer (7,34) dia-
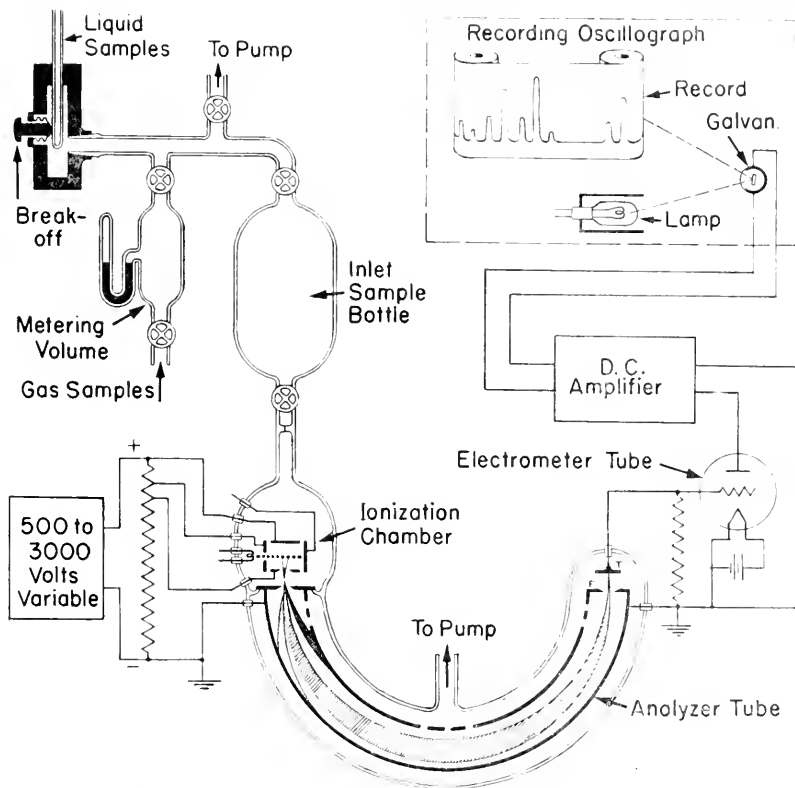
Fig. 3. Schematic diagram of the 180° Consolidated Engineering Company mass spectrometer (7).

grammed in Figure 3 and for others (5,12). Figure 3 is probably selfexplanatory but it appears deceptively simple since the numerous and complex electronic circuits are not included. Neither is the magnet, which must produce a uniform magnetic field sufficiently large to cover the entire region occupied by the tube and not merely a small central sector.

The technical skill required for determining isotope ratios with a mass spectrometer is not beyond that of most college-trained laboratory workers, assuming that the apparatus has been properly adjusted initially. Where the batteries shown in Figure 2 are replaced by electronic rectifying circuits, the services of someone skilled in electronics will be needed from time to time. Such a person should be available locally and on short notice if reasonably continuous operation is desired.

Complete installations of currently available mass spectrometers will probably cost in excess of ten thousand dollars (34,37,38). An instrument of the Nier type probably could be built in most university instrument shops for somewhat less. But it is almost necessary, even then, to have the complete ion tube assembly of Figure 2 made by someone who has had construction experience with them. The remainder of the apparatus consists of less critical items of either an electronic or glass-blowing nature. (See also 3a.)

4. Preparation of Samples for Analysis

Isotopic analyses of the complex organic substances that occur in nearly all biological studies must be made indirectly, since only gases or volatile liquids possessing suitable characteristics can be introduced into an ordinary mass spectrometer. Furthermore, it is often necessary to degrade a complex molecule stepwise so that an analysis can be made for isotopes occurring in the various positions or radicals. The final measurements, however, in all biological investigations to date have been carried out on simple gases.

The ultimate selection of a particular test gas rests, for each element, on several basic considerations. Low molecular weight is favored owing to the better spectral separation obtained. Molecules with a tendency to adhere to the inner walls of the evacuated spectrometer system are avoided since difficulty is experienced in removing them completely following each analysis. Water vapor falls in this category. Some substances decompose too readily upon contact with the hot filament; others furnish a varying yield of ionization products; still others produce ionized particles subject to interference from other particles of the same mass-to-charge ratio (Table II).

Recommended test gases for each of several isotopes, based on the considerations just outlined, will now be briefly discussed. For

deuterium, the hydrogen molecule itself has been found most convenient, even though the analyses are rather complicated (3, p. 99).

TABLE II

Mass Spectra of Some Ions[a] Usually Observable in Common Test Gases

Mass no.	H_2 98%	D_2 98%	N_2	O_2	CO_2	Uninvited particles
1	H^1	. . .	. . . . . .	. .	. . . . . .	. . . . .
2	H^1H^1	H^2	. . . . . .	. .	. . . . . .	. . . . .
3	H^1H^2	H^1H^2	. . . . . .	. .	. . . . . .	. . . . .
	$H^1[H^1]H^1$	. . .	. . . . . .	. .	. . . . . .	. . . . .
4	. . .	H^2H^2	. . . . . .	. .	. . . . . .	. . . . .
6	. . .	$H^2H^2H^2$	. . . . . .	. .	. . . . . .	. . . . .
12	. . .	. . .	. . . . . .	. .	C^{12}	. . . . .
13	. . .	. . .	. . . . . .	. .	C^{13}	. . . . .
14	. . .	. . .	N^{14}	. .	. . . . . .	. . . . .
	. . .	. . .	$(N^{14}N^{14})^{++}$	. .	. . . . . .	. . . . .
$14\frac{1}{2}$	. . .	. . .	$(N^{14}N^{15})^{++}$	. .	. . . . . .	. . . . .
15	. . .	. . .	N^{15}	. .	. . . . . .	. . . . .
	. . .	. . .	$(N^{15}N^{15})^{++}$	. .	. . . . . .	. . . . .
16	. . .	. . .	. . . . . .	O^{16}	O^{16}	O^{16}
17	. . .	. . .	. . . . . .	O^{17}	. . . . . .	$O^{16}\,H^1$
18	. . .	. . .	. . . . . .	O^{18}	. . . . . .	$H^1H^1O^{16}$
22	. . .	. . .	. . . . . .	. .	$(C^{12}O^{16}O^{16})^{++}$	
28	. . .	. . .	$N^{14}N^{14}$	. .	$C^{12}O^{16}$	$C^{12}O^{16}$
29	. . .	. . .	$N^{14}N^{15}$	. .	$C^{13}O^{16}$	CH_2NH?
	. . .	. . .	. . . . . .	. .	$C^{12}O^{17}$	. . . . .
30	. . .	. . .	$N^{15}N^{15}$	. .	$C^{12}O^{18}$	. . . . .
31	. . .	. . .	. . . . . .	. .	. . . . . .	CH_3NH_2
32	. . .	. . .	. . . . . .	$O^{16}O^{16}$	. . . . . .	. . . . .
33	. . .	. . .	. . . . . .	$O^{16}O^{17}$	. . . . . .	. . . . .
34	. . .	. . .	. . . . . .	$O^{16}O^{18}$	. . . . . .	. . . . .
36	. . .	. . .	. . . . . .	$O^{18}O^{18}$	. . . . . .	. . . . .
44	. . .	. . .	. . . . . .	. .	$C^{12}O^{16}O^{16}$	. . . . .
45	. . .	. . .	. . . . . .	. .	$C^{13}O^{16}O^{16}$	$(CH_3)_2NH$
	. . .	. . .	. . . . . .	. .	$C^{12}O^{16}O^{17}$	$C_2H_2NH_2$
46	. . .	. . .	. . . . . .	. .	$C^{12}O^{16}O^{18}$	. . . . .

[a] All ions carry a single positive charge unless indicated otherwise.

In very dilute samples of deuterium, the ions that occur in readily measurable concentrations are listed in the second column of Table II; in the very concentrated deuterium samples, in the third column.

For intermediate concentrations, the ionized particles appearing in both columns must be considered. Hydrogen samples for analysis are obtained by combustion of the biological compound (1, p. 56) and subsequent reduction of the resultant water is produced either by the use of a magnesium amalgam (20, p. 121) or by electrolysis (17, p. 130).

For oxygen isotopes, analyses can be made conveniently with molecular oxygen. The ions produced most plentifully are listed in the fifth column of Table II. Ions resulting from traces of water in the instrument interfere with atomic oxygen particles at mass positions 17 and 18. Consequently isotopic ratios are preferably based on the relative heights of the ion current peaks for masses 33 and 32, and for 34 and 32.

Carbon dioxide can also be used as the test gas for oxygen isotope analyses. Not all of the possible ions produced in carbon dioxide are shown in Table II. But in concentrated test samples one would observe peaks corresponding to all mass combinations of C^{12} and C^{13} with O^{16}, O^{17}, and O^{18} that are possible in a triatomic molecule. Where the O^{17} concentration is constant and the O^{18} is variable, abundance ratios can be based on measurements at mass numbers 44 and 46.

Determinations of N^{15} can be made fairly satisfactorily with the diatomic nitrogen molecule as the test gas. The detectable ionized particles that may be expected to appear will be found listed in Table II. The use of the atomic ions of masses 14 and 15 is counterindicated, owing to the presence of a peak in the $14\frac{1}{2}$ position. This results from doubly charged $N^{14}N^{15}$ ions and obviates the anticipated improvement in the spectral separation that would otherwise accrue from the lower masses of the atomic ions compared with the molecular ions. Other disadvantages of the atomic ions for ratio measurements have been discussed by Nier (1, p. 21). Accordingly, nitrogen analyses are normally based on measurements of the 28 and 29 ion current peaks. Some interference at the 28 and 29 positions is to be expected from atmospheric nitrogen contamination; also carbon monoxide and carbon dioxide must be excluded. In addition, Rittenberg (1, p. 32) cautions against still another impurity that can be present at mass 29 unless the Kjeldahl digestion is carried out according to definite, prescribed rules. To obtain nitrogen gas from biological organic compounds, the latter must undergo either an alkaline hydrolysis or a Kjeldahl combustion to yield an ammonium salt. Subjecting the

ammonium salt to a hypobromite treatment results in the liberation of nitrogen. Detailed procedures and precautions based on years of experience in isotope analysis have been published by Rittenberg (1, p. 31).

Carbon dioxide is the gas of choice for the mass spectrometric analysis of C^{13}. The numerous ionized particles produced are listed in Table II. Although the molecular ions have the greatest mass and therefore the least resolution, their peaks are measured because they possess the highest intensity. The partial interference at mass 45 owing to O^{17} can be corrected for satisfactorily. Were it not for the very low intensity of the atomic carbon peaks at the 12 and 13 positions, their improved resolution and freedom from interference would recommend them as a basis for analysis.

The preparation of carbon dioxide by the oxidation of organic compounds is quite convenient. Some workers employ persulfate oxidation (22). The wet combustion method of Van Slyke and Folch (21) has been used with success by several workers in the field (1, p. 43). A dry combustion process has also been used and a satisfactory apparatus and procedure have been described (1, p. 40). Decarboxylation methods of obtaining carbon dioxide from fatty acids and from amino acids (1, p. 41) should prove useful in special cases. The latter technique involves the use of ninhydrin (1,2,3-indanetrione hydrate)

5. Measurements and Maintenance

Before attempting measurements, the residual gas pressure in a mass spectrometer should be reduced to about 10^{-6} mm. of mercury. This may require heating the main evacuated part of the apparatus to perhaps 300°C. in a degassing operation. In fact, such baking may need to be repeated every time atmospheric air is admitted to the system. It is customary to maintain a vacuum continually in the apparatus by allowing the pumps to run without interruption. In that event, trap T_1 of Figure 2 must always be kept cold by the use of solid carbon dioxide or liquid air. During the progress of a measurement on a test sample the pressure in the body of the instrument may be approximately 2×10^{-6} mm. of mercury, but the gas pressure in the ionizing box will be still higher, perhaps 5×10^{-6} mm. of mercury.

Some hints on the practical maintenance of a mass spectrometer

have been furnished by Young (19), who also has analyzed the personnel and time factors involved in its full time operation.

In order to observe a peak corresponding to a given mass number, the electrical constants of the mass spectrometer must satisfy the following equation:

$$\frac{M}{e} \text{ (mass units)} = \frac{4.82 \times 10^{-5} \; r^2 (\text{cm.}) \; H^2 \, (\text{gauss})}{V \, (\text{volts})} \tag{2}$$

By "mass unit" we mean the ratio of the atomic or molecular mass to the number of unit charges on the ionized particle. For example, M/e for $C^{12}O^{16}O^{16}$ would be 44 for the singly charged ion, and 22 for the doubly ionized. The radius of curvature of the path around which the ions are deflected (15 cm. in the apparatus of Figure 2) is
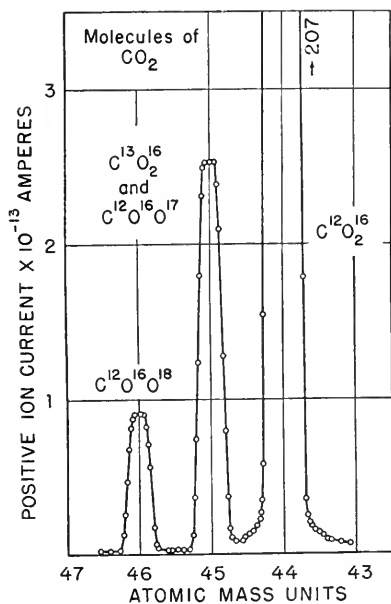
Fig. 4. Mass spectrum of carbon dioxide from early data of Nier (4), showing only the most abundant molecular ions (see Table II). Ion-accelerating potential was approximately 500 v. Note the resolution, the flat-topped peaks, and the background current.

represented by r, the uniform magnetic field by H, and the ion-accelerating potential by V. In practice it is usually convenient to maintain a constant magnetic field and to vary V. When observing particles with mass numbers as different as hydrogen and carbon dioxide, for example, it may also be necessary to vary H. To calibrate a mass spectrometer initially, fix either H or V at a reasonable value and then

compute the proper setting of the other variable for the selected particle by means of equation (2). The final searching, of course, is done experimentally. Peaks corresponding to water vapor and carbon monoxide are helpful here; they can always be detected even with thoroughly degassed apparatus.

When measurements of isotopic ratios are made with a nonrecording mass spectrometer, the plotted observations have an appearance similar to the experimental curve for a carbon dioxide analysis reproduced in Figure 4. In practice, observations are required on the top of the peaks only, except for just enough readings elsewhere to enable corrections to be made for the background. To avoid errors that may arise from any instability in the electrical circuits or from gas pressure variation in the ion source, a complete set of data should include perhaps a dozen observations. These should be taken alternately on each of the peaks and within the shortest time possible. Hence the electrical circuits should permit an almost instantaneous shift from one peak to another (see 5). Multiple sets of control equipment, one for each peak, are required in which provision has been made for adjusting not only the proper ion-accelerating voltage corresponding to each peak, but also the correct sensitivity of amplifier H (see Fig. 2). The latter may involve shifting from one input resistor, R, to another.

6. Calculation of Results

Typical experimental observations for a C^{13} determination with a carbon dioxide test sample are shown graphically in Figure 4. Ion current readings cover mass numbers 44, 45, and 46. The noticeable background deflection that appears in this figure must be subtracted from the various peak heights before their ratios are computed. Further, the 45 peak has received a small contribution to its height from the presence of $C^{12}O^{16}O^{17}$ molecules in addition to the desired $C^{13}O^{16}O^{16}$ molecules. This constant O^{17} correction amounts to 0.08% and may be made either before or after the calculation of the 44/45 ratio. Assuming that the two corrections just mentioned have been made and that the ratio, $^{44}_{45}R$, of the resultant ion currents at mass 44 to that at mass 45 has been calculated, then the concentration of C^{13} in atom per cent is given by:

$$C^{13} = \frac{(C^{13})100}{(C^{12}) + (C^{13})} = \frac{100}{^{44}_{45}R + 1} \text{ atom per cent} \tag{3}$$

Whenever an isotopic element being studied is represented by a single atom in the ionized test particle, even though one or more other elements may also be present, the concentration of the element under test can be calculated usually by means of an appropriate substitution in equation (3). Of the common stable isotopes encountered in biological studies, the measurement of sulfur, either as sulfur dioxide or as hydrogen sulfide, furnishes a second example.

If the isotopic element being measured occurs twice or more in the test molecule, the calculations are somewhat more involved. For example, when N^{15} determinations are made using the diatomic nitrogen molecule, probability considerations indicate the following molecular configurations: $N^{14}N^{14}$, $N^{14}N^{15}$, $N^{15}N^{14}$, and $N^{15}N^{15}$. If the concentration of N^{15} in atom per cent equals P, the complementary atomic concentration of N^{14} is $(100 - P)$. The relative molecular concentrations at equilibrium of the three molecular species with masses 28, 29, and 30 will be $(100 - P)^2$, $2P(100 - P)$, and P^2, respectively. Since the ionization current readings of the mass spectrometer will be proportional only to the concentrations of the molecular ions having masses 28, 29, and 30, the ratio of the corrected peak readings for masses 28 and 29 will be expressed as follows:

$$_{29}^{28}R = \frac{\text{mass 28 peak}}{\text{mass 29 peak}} = \frac{(100 - P)^2}{2P(100 - P)} = \frac{100 - P}{2P} \tag{4}$$

The solution of this equation indicates that the concentration, P, of N^{15} can be calculated from measurements at only two mass numbers. That is:

$$P = \frac{100}{2\,_{29}^{28}R + 1} \text{ atom per cent} \tag{5}$$

When the height of the peak at mass 30 indicates a sufficiently high concentration, the calculation can be based also on either of the other two ratios, as indicated by the following equations:

$$_{30}^{29}R = \frac{2P(100 - P)}{P^2} \text{ or } P = \frac{200}{_{30}^{29}R + 2} \text{ atom per cent} \tag{6}$$

$$_{30}^{28}R = \frac{(100 - P)^2}{P^2} \text{ or } P = \frac{100}{_{30}^{28}R + 1} \text{ atom per cent} \tag{7}$$

Determinations of P by two independent methods serve as a valuable check on the equilibrium attained by the ionic test particles and help

to eliminate accidental mistakes in both measurements and calculations.

With molecular oxygen as the test gas, determinations of low concentrations of O^{17} and O^{18} could be calculated on the basis of measurements on only two peaks by modifying equation (5) as follows:

$$P \text{ (of } O^{17}) = \frac{100}{2 \, _{33}^{32}R + 1} \text{ atom per cent}$$

and:

$$P \text{ (of } O^{18}) = \frac{100}{2 \, _{34}^{32}R + 1} \text{ atom per cent}$$

When O^{18} is used as a tracer and carbon dioxide is the test gas, its concentration may be found from:

$$P \text{ (of } O^{18}) = \frac{100}{2 \, _{46}^{44}R + 1} \text{ atom per cent}$$

since the contribution to the height of the 46 peak owing to $C^{13}O^{16}O^{17}$ is a negligible factor.

Before calculating low deuterium concentrations by an analogous formula based on masses 2 and 3, a preliminary correction of the mass 3 reading owing to the presence of triatomic hydrogen molecules is considered essential (see 3, p. 99). Intermediate concentrations of hydrogen and deuterium offer additional difficulties owing to the presence of several types of ion at both the mass 3 and 2 positions.

The general equation for the concentration, P^m, of an isotopic element, which may contribute two atoms with masses of either m or n to the charged particle being measured, is:

$$P^m = \frac{[\frac{1}{2}P^{(n+m+s)} + P^{(2m+s)}]100}{P^{(2n+s)} + P^{(n+m+s)} + P^{(2m+s)}} \text{ atom per cent} \qquad (8)$$

where each P on the right of the equation represents the measured ion current corresponding to the mass numbers given by the exponents; s is the sum of the atomic mass numbers for any foreign atoms common to all of the ionic particles. This equation therefore requires measurements for all three mass numbers. For example, it is directly applicable to determinations of N^{15} for intermediate concentrations.

Measurements of N^{15}, and also of the oxygen isotopes, often must be corrected for the presence of air that has leaked into the mass

spectrometer (*1*, p. 38). Since it is practically impossible to avoid having some air in the instrument, such corrections are usually indicated. When N^{15} is being measured, air is shown to be present by the occurrence of an oxygen peak at mass 32 and confirmed if an argon peak appears at mass 40. Although the ratio of nitrogen to oxygen in the atmosphere is approximately four, the air nitrogen contribution to the mass 28 peak by a contaminated test sample is usually greater than would be indicated by multiplying the height of the mass 32 peak by four. The ratio as actually measured depends on instrumental and other factors, including differential ionization probabilities for the two elements. A proper correction factor, characteristic for each mass spectrometer, must be determined so that corrections can be applied to the readings at the various mass numbers, should air leaks be encountered.

Instead of expressing isotope concentrations in atom per cent as calculated by the foregoing equations, there is an increasing tendency to publish results in terms of the atom per cent *excess* concentration as compared to the normal abundance or to a laboratory standard. These excess values are computed by subtracting either the normal concentration, as tabulated in Table I, or the concentration of an arbitrary, but permanent, laboratory standard. The latter procedure is preferable for several reasons, including a continual check on the performance of the mass spectrometer and on any variations in the concentration of the normal isotope. Some types of calculations, as used for example with the isotope dilution technique, are simplified when concentrations are expressed in terms of atom per cent excess. Since arbitrary standards of some elements prepared from chemical salts may possess higher concentrations than occur normally in biological preparations, negative values will be encountered occasionally (*1*, p. 28).

D. LIMITATIONS AND SOURCES OF ERROR

1. Absolute Accuracy of Isotope Concentrations

To achieve high absolute accuracy in stable isotope measurements with a mass spectrometer is very difficult. The sources of the errors are inherent in the design of the instrument itself and cannot be eliminated easily. Fortunately most of the biological applications of stable isotopes are based on a determination of their relative abun-

dance only, so that precise measurement of their absolute concentration is not essential.

Since the rate of diffusion of gas molecules depends on their mass, it is to be expected that the lighter of two isotopic compounds would enter the spectrometer more readily than the heavier. Unless counterbalanced, this preferential effect would result in a relatively higher reading for the concentration of the lighter isotope. As a matter of fact, a similar but compensatory effect does occur in the diffusion pump, which is constantly evacuating the spectrometer, thus tending to reduce the observed concentration of the lighter isotope component. Hence these two factors tend to counterbalance each other and the net error introduced is probably not large. Discrimination between particles of different masses may also occur in the ion source of a mass spectrometer (2, p. 539). As a consequence the concentration ratios may be affected even to the extent of a few per cent; the effect is greater for higher mass ratios. While such discrimination can be reduced by employing an ion source with a more complicated design, it may not be profitable to do so in the routine analytical spectrometer since an accurate absolute measurement of the concentration is not necessary in tracer experiments.

A number of other sources of error that affect the accuracy of the absolute abundance ratio will be discussed in the sections that follow, since they also influence the relative accuracy of isotope determinations (see also 3, p. 101).

2. Relative Accuracy

Mass spectrometric analyses of biological materials usually comprise comparisons between the concentration of an element used as a tracer and that of its naturally more abundant isotope. As pointed out in preceding paragraphs, this ratio may not be correct in an absolute sense to within a few per cent. However, if it is used only with reference to similar ratios all of which are based on the same standard gas sample, as discussed later (Sect. D8), it yields entirely adequate information for use in tracer studies. Thus, the relative accuracy of a pair of ratio determinations may be independent of a number of sources of error that affect adversely the absolute determination of either one. Still, the relative ratio does depend on the quality of the equipment and on the technical skill of the operator in a number of ways. These will now be briefly discussed.

To secure accurate determinations of even relative ratios, several conditions must be maintained constant during the measurements. One important factor concerns the pressure of the gas in the neighborhood of the ion source. Adequate control of this pressure requires the use of a mechanism that permits the small gas sample to leak extremely slowly into the ionizing compartment of the spectrometer, yet without any time variation. To satisfy this requirement, the initial sample of gas must be sufficient to enable an adequate and constant differential pressure to be maintained across the leak mechanism (L in Fig. 2) throughout the course of the ratio determinations. Likewise the pumps, which are continually evacuating the mass spectrometer, must operate uniformly during the measurement period if pressure fluctuations of the test gas within the ion source are to be avoided.

In order to create a constant supply of ions, it is also necessary to control accurately the emission of electrons from the filament with respect to their quantity, speed, and direction. This involves not only the regulation of the filament temperature and the electron-accelerating voltage applied between the filament and the electron collector, but also the strength of the magnetic field, which collimates the stream of electrons. With available electronic regulators, the variations in the magnitude of each of these quantities can and should be maintained within less than 1% during the course of an analysis.

When the current through the principal magnet is stabilized at some constant value, the emergence from the exit slit of particles having a particular ratio of mass to charge is determined by the magnitude of the ion-accelerating voltage. It will be noted in Figure 4 that the ion current peaks corresponding to various masses possess flat tops. The intentional presence of these flat tops makes the peak readings less sensitive to slight variations in either the ion-accelerating voltage or the current through the principal magnet. However, a stabilizer will still be necessary if the equipment operates from an alternating voltage supply.

The effect of transient fluctuations resulting from instability of the ion current amplifier or other electrical circuit components, as well as from the causes just enumerated, can be minimized statistically by taking an extended series of measurements on the same sample, with alternate readings on the heavier and lighter isotope components. Frequent comparisons with the standard sample serve a similar purpose.

The different mass peaks in the ionization current curve of Figure 4 must be sufficiently well separated to avoid appreciable interference with each other. This separation is a function of the slit widths of a mass spectrometer, but it decreases with increasing mass of the ions. For cases of insufficient resolution, and particularly where peaks differ markedly in intensity, an estimated correction may need to be applied to the intensity of the weaker peak, thus decreasing the accuracy.

3. Natural Variations in Isotope Concentrations

Generally accepted values for the natural concentrations of the isotopes considered in detail in this chapter are listed in Table I. Agreement between published values for some of these absolute ratios has not been very good. The small discrepancies probably result largely from differences in particle discrimination among the various mass spectrometers. In addition, it has been demonstrated that there is a natural variation in the concentration of isotopes depending on their past history (2, p. 561; 17, p. 8).

"Normal" C^{13} concentrations have been found to vary as much as 5% between carbon obtained from limestone and carbon obtained from plant cells. Similarly O^{18} from carbonate ores has a concentration about 4% higher than that found in water generally, but deuterium from widely different sources is constant to within a few thousandths of 1%. There has been little work on variations in the natural occurrence of N^{15}, but some analyses made on amino acids isolated from the protein of mammals have indicated a slight excess varying from 0.000 to 0.006 atom per cent N^{15} over that found in air. In one study with arginine, an excess of 0.008 atom per cent N^{15} was found. No information concerning natural concentration variations for the several isotopes of sulfur is known to the writer. It may well be expected that very small variations in concentration will occur in some biochemical reactions. While a systematic study of such variations could be of interest in itself, perhaps its most useful function would consist of indicating probable limits of accuracy for investigations employing isotopes as tracers. Most past experiments have had such a large margin of safety with respect to concentration changes that the conclusions have not been jeopardized owing to these slight natural isotopic variations. While selective behavior of living organisms toward low concentrations of isotopic compounds is ab-

sent or negligible in most cases, this is not true for high concentrations
of deuterium (24, p. 220).   In fact, deuterium as a tracer should not
be employed in concentrations higher than about 10% unless its selec-
tive behavior is also investigated.   Heavy water may be definitely
toxic to organisms when its deuterium content exceeds 25%.

4. Minimum Size of Test Samples

The minimum quantity of gas required for a mass spectrometer
test sample depends somewhat on the particular instrument.   Two
frequently stated estimates of a satisfactory volume are either 100
ml. of a gas at 1 mm. of mercury pressure or 0.1 ml. of gas at a pres-
sure of 1 atmosphere.   Using either of these criteria for carbon di-
oxide, one arrives at a weight of approximately 0.2 mg.   Expressed
in terms of carbon content only, something less than 0.1 mg. should
be sufficient for an analysis.   On the same basis hydrogen would re-
quire approximately 0.01 mg. for a minimum test sample.   In work-
ing with such small quantities of matter, it is usually necessary to
allow greater tolerances on reaction yields than would otherwise be
considered normal.   For example, organic samples to be tested for
C^{13} or N^{15} might well contain approximately 1 mg. of either carbon
or nitrogen previous to combustion.

5. Limiting Dilution Factors

The factor by which an isotope having a specified original concen-
tration can be diluted and still be detected reliably depends both on
the accuracy with which concentrations can be measured and on the
natural abundance of the isotope.   The permissible dilution is
greater for isotopes whose natural abundance is smaller.   Where
ratio determinations can be made with an accuracy of 1% and where
initial concentrations of the isotope are in the neighborhood of 50
atom per cent excess, N^{15} can be detected following a 10,000-fold
dilution.   An equivalent value for S^{34} is 1000-fold, for C^{13} 5000-fold,
and for deuterium 5000-fold (2, p. 563).   These dilution factors are
more than adequate for numerous types of investigation.   They are
not as large, however, as can be obtained by the use of artificial radio-
active isotopes, which are greatly favored in this respect because their
natural concentration is zero.   Since, with extreme care, higher ac-
curacy can be obtained than assumed above (1, p. 29), some improve-
ment over the ratio values cited is possible.   Most studies to date
have not even approached the above dilution limits.

6. Exchange Reactions

The application of stable isotopes to biological problems is limited markedly by the ease with which a number of elements exchange positions between or within molecules.   This type of behavior has been known for a long time in the case of some elements and suspected for others.   Owing to the fact that this exchange is not usually observable in ordinary chemical reactions, it had not received much attention prior to the availability of the isotope method for studying reaction rates.

The exchange of hydrogen atoms between the molecules of a solute and a solvent is a most rapid process for many chemical radicals (*15*, p. 62).   This unstable condition obtains for practically all polar groups, but special mention must be made of the carboxyl, hydroxyl, and amino radicals that are so common in biochemical systems.   As a consequence, deuterium introduced into these labile groups cannot be used as a tracer.   In general, deuterium attached directly to carbon atoms has been found to be stably bound, particularly in methyl and methylene radicals.   There are exceptions, however, as in the case of hydrogen atoms attached to carbon atoms adjacent to keto groups, since such hydrogen atoms are known to undergo enolization.   The conclusion reached, at least with respect to the use of deuterium as a tracer is that its stability with reference to any proposed chemical system must first be confirmed.   There has been no evidence to the effect that an isotope that has been shown to be stably bound with reference to a given chemical system *in vitro,* is not stable *in vivo* also. Not only the stability of the initial compound to be synthesized but also of any subsequent compounds to be isolated and analyzed for the presence of deuterium must be checked.

The ready exchange of oxygen atoms in solution (*15*) has made the O^{17} and O^{18} isotopes practically valueless in biological tracer work. Oxygen exchange in carbonyl groups is particularly rapid, thus eliminating studies involving aldehydes and ketones.   The oxygen in carboxyl groups plays a variable exchange role, depending on the compound and on the environmental conditions, so that individual confirmations of its behavior under each given set of circumstances would be indicated.   The oxygen in hydroxyl groups is stable in many common alcohols, at least under mild acid and alkaline conditions, and may be found useful as a tracer.   The same seems to apply to the hydroxyl groups in pentoses and hexoses but not to the carbonyl

oxygen. Existing exchange evidence indicates that the use of oxygen isotopes in the sulfate radical and in the nitro group may be permitted under some circumstances, but the former does exchange with water in concentrated acid solution.

Exchange reactions of the sulfur present in various inorganic radicals has also been studied (cf. 15).

Nitrogen atoms attached to carbon appear to be molecularly stable in biological systems as far as the amino acids are concerned. In view of their great importance, this is very fortunate. Urea nitrogen under some circumstances may undergo very slow exchange in solution. Gaseous nitrogen has been shown not to exchange with molecularly combined nitrogen in solutions of either sodium nitrate or sodium nitrite.

Carbon has been found to exchange rapidly between the carboxyl group of sodium acetate and acetic anhydride (17, p. 159). In general, one would expect stability of carbon-to-carbon bonds, but it should not be assumed without experimental confirmation.

7. Interfering Particles

Two types of interference are recognized. One type results when the test gas employed contains one or more chemical elements in addition to the specific one under investigation. The isotopic constituents of any accessory elements bring about the interference. An example occurs in the use of carbon dioxide for the analysis of C^{13}. The peak at mass 45 results not only from $C^{13}O^{16}O^{16}$ but also from the presence of $C^{12}O^{16}O^{17}$ ions (see Fig. 4). Thus the accessory oxygen atoms with their possible masses of 16, 17, and 18 give rise to higher measured ratios than can be ascribed to the excess concentration of C^{13} alone. Corrections for this type of interference can be made readily if the concentrations of the offending atoms are known; an illustration of the calculations involved has been given earlier.

A second type of interference can be ascribed to the presence of radicals of the test gas. These may be either singly or multiply ionized. Reference may be made to the methane spectrum in Figure 1 as an illustration. The peak at mass 15 owing to CH_3 radicals is almost as high as the CH_4 peak. Consequently in any analysis for C^{13}, nearly half the C^{13} atoms would contribute to the mass 16 peak as $C^{13}H_3$ ions, together with the $C^{12}H_4$ ions, rather than appear at

mass 17 in the form of $C^{13}H_4$. Corrections for this type of interference are often quite difficult, but even here satisfactory analyses can be made provided the offending radical is recognized. First, standardization measurements are made covering each of the various compounds present and also their disintegration products under carefully controlled conditions of operation (*18*). Of primary importance is the energy of the bombarding electrons that produce the radicals and bring about their ionization. Corrections for the contributions of numerous molecules and radicals to a given mass peak are considered routine when mass spectrometers are used for quantitative analyses of organic mixtures such as are encountered in the petroleum industry (*7*).

It may be well also to mention here the type of self-interference that may result from leaks in the apparatus. Perhaps the most annoying illustration occurs in the use of molecular nitrogen for N^{15} measurements; any leak of atmospheric air into the apparatus is troublesome. The same is true when isotopic oxygen determinations are made with molecular oxygen.

8. Standard Samples for Comparison

As already pointed out, it is very difficult and quite unnecessary to determine *absolute* values of concentration of the stable isotopes with great accuracy. Biological requirements are amply satisfied provided accurate relative determinations of isotope ratios can be obtained. However, it is highly desirable always to relate such measurements to some permanent standard sample.

For a standard sample in the case of C^{13}, an adequate reserve of a carbonate salt should be set aside. For each run, a standard test sample is admitted to the mass spectrometer so that all calculations of the isotope concentrations in terms of excess per cent can be based on the same standard. This procedure serves as a check on the operation of the mass spectrometer too, not only during each set of measurements but also from day to day and from month to month. With equipment that is functioning satisfactorily, daily variations in the recorded ratios for the standard sample should lie within rather narrow limits. The preservation of a permanent laboratory standard will permit later comparisons with other laboratories should the occasion arise. Perhaps eventually all data could be related to the same national or international standard.

9. Over-All Accuracy

For the type of biological investigation we have been discussing, the accuracy obtainable in routine determinations of isotopic concentration is usually quite adequate. There seems to be universal agreement among the workers in this field that such determinations can be made readily within an over-all accuracy limit of 1%. There is reason to believe that carefully conducted measurements can reduce this limit by a factor of two, or perhaps, with extreme care, by a factor of four. It is very important to realize the limits of accuracy of an experimental method, even though an investigation may not necessitate pushing a method to its ultimate limit. Wherever possible, it is preferable to operate at a comfortably safe distance inside the extreme limits of accuracy. Although the significance of experimental results in each case must be determined eventually by an adequate statistical evaluation of the data, one would prefer to know in advance that there existed a sufficient margin of safety so that the experimental conclusions would not be open to question.

E. ACCOMPLISHMENTS WITH INDIVIDUAL ELEMENTS

1. Deuterium and Heavy Water

Since heavy hydrogen was the first stable isotope to be isolated in quantity, a voluminous literature on its biological applications is to be expected. The behavior of numerous organisms subjected to heavy water of various concentrations was observed in early experiments and the deuterium label was traced to a great variety of body constituents (24). Much useful information is contained in papers presented at a 1938 symposium on deuterium (15). A very important group of researches using deuterium as a label concerns fatty acid metabolism and the general behavior of fats in the animal body. It happens that deuterium may be stably bound to carbon in the higher fatty acids and by rather simple synthetic methods (see 17, p. 139). Two excellent résumés of this work by the principal investigators (23,24) stress the complex relationships and interconversions continually in progress among this important group of body constituents. The deuterium label has also been used with several amino acids (24). The transmethylation reaction involving methionine deserves special mention; so does the discovery of the inert nature of lysine, which is

exceptional in that it does not react with deuterium *in vivo*. An extensive bibliography has recently been published (*31a*).

2. Carbon

Recent isotopic studies in the field of intermediary carbon metabolism have emphasized the fixation of carbon dioxide, a reaction all but inaccessible to other methods of approach. The part played by C^{13} in the elucidation of the mechanism of carbon dioxide fixation has been presented in an excellent review by Wood (*28*). Isotopic carbon investigations of metabolic mechanisms in general, including fats, have been reviewed at length recently by several workers (*22a*; *27*; see also *3*). The transport of nutrients as well as the products of photosynthesis has been investigated in plants with C^{13}; a guide to these studies occurs in an article by Rabideau and Burr (*26*). Only a beginning has been made in attacking the problem of photosynthetic reaction mechanisms and the chemical nature of the intermediates (*22a*; *27*, pp. 121–122).

3. Nitrogen

The important role of nitrogen in amino acid and protein metabolism has been actively investigated with N^{15} ever since this isotope was made available. Most of the work has been carefully reviewed from time to time. The early studies are covered in articles by Schoenheimer, both alone (*23*) and in collaboration with Rittenberg (*24*). Later reviews have also appeared (*29*). Nitrogen fixation by plants has also been reviewed recently (*31*). Satisfactory contact with current progress in the field can be maintained only by consulting the most recent annual review or the expanding literature in the journals. The fact that only a brief space is devoted here to directing attention to N^{15} studies on the translocation and intermediary metabolism of proteins, amino acids, and related compounds should not be construed as an appraisal of their value. The results are regarded generally as being of tremendous importance.

4. Oxygen

Apart from numerous studies on oxygen exchange reactions of organic compounds (*15*), few biological applications of oxygen isotopes have been made. Too rapid exchange is the principal reason. The

direct conversion of oxygen to carbon dioxide in the lungs has been shown experimentally, and O^{18} has been employed in an effort to determine the source of oxygen in the photosynthetic reaction (see *2*, p. 568).

5. Sulfur

Stably bound sulfur isotopes have been employed to follow the transport of sulfur compounds in plants (*25*). The synthesis and metabolic behavior of methionine have been the object of several investigations, both with radioactive S^{35} (*17*, p. 205) and with the stable S^{34} (*27*, p. 151). In fact both C^{13} and S^{34} labels were used to mark methionine. Cystine has also been synthesized and studied with S^{35} (*30*).

References

GENERAL

1. Symposium: *Preparation and Measurement of Isotopic Tracers*. Edwards, Ann Arbor, 1946.

2. Symposium: The Mass Spectrometer as an Industrial Tool. *J. Applied Phys.*, **13**, 525 (1942).

3. Symposium: *The Use of Isotopes in Biology and Medicine*. Univ. Wisconsin Press, Madison, 1948.

3a. Stewart, D. W., "Mass Spectrometry," in *Physical Methods of Organic Chemistry*, 2nd ed., A. Weissberger, ed., Interscience, New York, 1949, Chap. XXXI.

3b. Hevesy, G., *Radioactive Indicators*. Interscience, New York, 1948.

MASS SPECTROMETERS

4. Nier, A. O., *Rev. Sci. Instruments*, **11**, 212 (1940); **18**, 398 (1947) (construction details of 60° model).

5. Brown, H., J. J. Mitchell, and R. D. Fowler, *Rev. Sci. Instruments*, **12**, 435, 556 (1941) (180° type; electronic control circuits).

6. Taylor, J. E., *Rev. Sci. Instruments*, **15**, 1 (1944) (60° permanent magnet model; C^{13} separation).

7. Washburn, H. W., H. F. Wiley, S. M. Rock, and C. E. Berry, *Ind. Eng. Chem., Anal. Ed.*, **17**, 74 (1945) (180° instrument with automatic recording).

8. Nier, A. O., C. M. Stevens, A. Hustrulid, and T. A. Abbott, *J. Applied Phys.*, **18**, 30 (1947) (60° leak detection type; permanent magnet).

ACCESSORY EQUIPMENT

 9   Penick, D. B., *Rev. Sci. Instruments*, **6,** 115 (1935) (electrometer tube, d.c. amplifier).

10.   Ridenour, L. N., *Rev. Sci. Instruments*, **12,** 134 (1941) ("Magic Eye" ionization gage).

11.   Taylor, R. C., and W. S. Young, *Ind. Eng. Chem., Anal. Ed.*, **17,** 811 (1945) (sintered glass leak).

12.   Thomas, H. A., T. W. Williams, and J. A. Hippel, *Rev. Sci. Instruments*, **17,** 369 (1946) (mass spectrometer type of leak detector; ion gage).

13.   Nier, A. O., E. P. Ney, and M. G. Inghram, *Rev. Sci. Instruments*, **18,** 191, 294 (1947); *cf. ibid.*, **18,** 18 (1947) (adjustable gas leak and inverse feedback d.c. amplifier; electrometer tubes).

14.   Hunt, F. V., and R. W. Hickman, *Rev. Sci. Instruments*, **10,** 6 (1939) (high voltage regulated power supply).

METHODS AND TECHNIQUE

15.   Reitz, O., *Z. Elektrochem.*, **45,** 100 (1939) (oxygen exchange reactions). For deuterium, see *ibid.*, **44,** 1 (1938);   for sulfur, see *Chem. Revs.*, **27,** 255 (1940).

16.   Rittenberg, D., and G. L. Foster, *J. Biol. Chem.*, **133,** 737 (1940) (isotope dilution technique;   *cf. Science,* **107,** 228 (1948).

17.   Kamen, M. D., *Radioactive Tracers in Biology.*   Academic Press, New York, 1947 (methods, applications;   syntheses with C, H, and S).

18.   Brewer, A. K., and V. H. Dibeler, *J. Research Natl. Bur. Standards*, **35,** 125 (1945) (hydrocarbon gas analysis).

19.   Young, W. S., *Natl. Petroleum News* (Technical Section), March 6, 1946 (operation and maintenance hints).

20.   Henriques, F. C., Jr., and C. Margnetti, *Ind. Eng. Chem., Anal. Ed.*, **18,** 420 (1946) (microdecomposition of water).

21.   Van Slyke, D. D., and J. Folch, *J. Biol. Chem.*, **136,** 509 (1940) (carbon combustion technique).

22.   Osburn, O. L., and C. H. Werkman, *Ind. Eng. Chem., Anal. Ed.*, **4,** 421 (1932) (persulfate oxidation of carbon).

22a. Calvin, M., C. Heidelberger, J. C. Reid, B. M. Tolbert, and P. E. Yankwich, *Isotopic Carbon.*   Wiley, New York, 1949.

22b. Radin, N. S., *Nucleonics.* **1,** No. 2, 49 (1947).

BIOLOGICAL ACCOMPLISHMENTS

23.   Schoenheimer, R., *Dynamic State of Body Constituents.*   Harvard Univ. Press, Cambridge, Mass., 1942 (fat and protein metabolism, D, and N^{15}).

24.   Schoenheimer, R., and D. Rittenberg, *Physiol. Revs.*, **20,** 218 (1940) (animal metabolism;   methods using D and N^{15}).

24a. Bigeleisen, J., *Science*, **110**, 14 (1949).

25. Thomas, M. V., R. H. Hendricks, L. C. Bryner, and G. R. Hill, *Plant Physiol.*, **19**, 227 (1944) (sulfur transport in plants).

26. Rabideau, G. S., and G. O. Burr, *Am. J. Botany*, **32**, 349 (1945) (C^{13} transport studies in plants).

27. Buchanan, J. M., and A. B. Hastings, *Physiol. Revs.*, **26**, 120 (1946) (C^{13} in intermediary metabolism).

28. Wood, H. G., *Physiol. Revs.*, **26**, 198 (1946) (C^{13} in intermediary metabolism).

29. Rittenberg, D., and D. Shemin, *Ann. Rev. Biochem.*, **15,** 247 (1946) (metabolism of proteins and amino acids; also see later volumes, especially **16,** 1947).

30. Melchior, J. B., and H. Tarver, *Arch. Biochem.*, **12**, 301, 309 (1947).

31. Wilson, and R. H. Burris, *Bact. Revs.*, **11**, 41 (1947); see also *Ann. Rev. Biochem.*, **14**, 685 (1945).

31a. Kimball, A. H., ed., *Bibliography of Research on Heavy Hydrogen Compounds.* McGraw-Hill, New York, in press (2000 references).

COMMERCIAL SOURCES

32. Stuart Oxygen Co., San Francisco, Calif.

33. Eastman Kodak Co., Rochester, N. Y.

34. Consolidated Engineering Co., Pasadena, Calif.

35. Norsk Hydro-Elektrisk Co., Kvaelstofaktieselskab, Oslo, Norway.

36. Isotopes Division, U. S. Atomic Energy Commission, Oak Ridge, Tenn.

37. General Electric Co., Schenectady, N. Y.

38. Process and Instruments, Brooklyn 22, N. Y.

RADIOACTIVE TRACERS

ADOLF F. VOIGT, *Institute for Atomic Research and Iowa State College*

A. INTRODUCTION

The use of radioactive tracers in biological research was begun a quarter of a century ago in the research of Hevesy (*11*), who studied

the uptake of lead by plants using a naturally occurring radioisotope of lead.   During the past 25 years the use of this technique has increased immeasurably.   The growth was slow at first since the early work was confined to the elements at the bottom of the periodic table for which radioisotopes occur naturally and the biological interest in these elements is not overwhelming.   The early 1930's was a period of numerous discoveries of fundamental importance in physics.   In addition to the discovery of the neutron and the positron and the development of "atom-splitting" ion accelerators, it was found that radioactivity was not limited to the heavy elements, but that radioactive isotopes could be prepared from every element in the periodic table.   The newly discovered neutron proved to be one of the most successful tools for the production of these artificial radioactive isotopes and the ion accelerators, in particular the cyclotron, were soon applied to the production problem both by direct action and as the most powerful neutron producers then known.

It is not surprising that the number of workers using these isotopes in biological research increased rapidly as soon as the isotopes became available.   A review written in 1940 contains a hundred references to work published prior to that time (12).   Every institution that had a cyclotron also had active research groups in biological applications of the tracers it produced, but it was somewhat more difficult for people outside these fortunate institutions to obtain tracers.   With the end of the war a small part of the nuclear energy reactors was turned over to the manufacture of radioisotopes for general tracer research use. In June, 1946, the Manhattan District, which operated the atomic energy project, announced the availability for sale to accredited research institutions of a large number of radioactive isotopes (13). This met with vociferous approval by the scientists of the country who felt that this technique could be used to solve many perplexing problems.   The project as continued by the Atomic Energy Commission has been eminently successful.   In the first year of its operation a thousand shipments of radioisotopes were made (14).

With the incomparably more available and less expensive radioisotopes that the nuclear "pile" supplies, the urge to do tracer research spreads out from a few to almost all research institutions.   The reduced cost puts this approach within the reach of many colleges whose budget could never have afforded isotopes previously.   It is reasonable then that many people will wish to do tracer research who have not had personal contact with other men in the field from whom they

could learn the techniques involved. It is hoped that this chapter will serve as an introduction to the method and to its rather extensive literature for such persons.

B. TERMINOLOGY OF RADIOACTIVITY

In the use of radioactive tracers there are certain ideas and terms of such basic importance that some time should be spent in discussing them before proceeding to the methods of applying tracers to biological problems. For further information see one of the books on this subject (4-10). The term isotope itself has been used with several different meanings and for clarification a new term "nuclide" had been introduced by Kohman (15) to mean any particular nuclear species, with given atomic number, Z, and mass number, A. The term isotope is then reserved for its original meaning, that of nuclides with the same Z and different A. Other terms frequently met are isobars, nuclides with different Z but the same A, and isomer (in the radiochemical sense), meaning "with the same Z and A but having different nuclear properties."

1. Types of Radiation and Decay Processes

Radioactive atoms are characterized by the emission of ionizing radiation. This ionization is capable of being extremely harmful in large doses, but the point at which the radiation may affect the biological system is not well defined. Since, in tracer studies, it is necessary to avoid affecting the system by the radiation, a consideration of this subject is essential in planning the experiments. The relative effects of the various radiations is considered briefly below along with their other properties, but a much more complete discussion of some of these points is given in Chapter XV.

The radiations by the emission of which radioactive elements decay are: α particles or helium nuclei; β particles, which are electrons if negative in sign and positrons if positive; and γ rays, which are similar to X rays.

α rays are highly energetic, most of them carrying an energy of 4-6 m.e.v. (One million electron volts, 1 m.e.v., is that energy an electron has after being accelerated by 10^6 v. It is equivalent to 1.6 $\times 10^{-6}$ erg.) The "range" or maximum penetration of these particles is 3-5 cm. of air or roughly 30-50 μ of tissue. Since α-emitting

isotopes are confined to the elements with atomic numbers of 83 and greater, they are not of wide interest as biological tracers.

Negative β rays are electrons ejected from the nucleus. A nuclide with a greater number of neutrons than its stable isotopes tends to transform itself, by emitting negative β rays, into its isobar of the element of next higher atomic number, corresponding to the hypothetical process:

$$_0n^1 \longrightarrow {}_1H^1 + {}_{-1}\beta^0$$
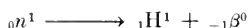
or conversion of a neutron into a proton. In general the energy of β rays is considerably less than that of the α particles, those from the useful tracers ranging from 0.05 to 2 m.e.v. The β rays from a given substance do not all have the same energy but have an energy distribution, whose maximum value is usually listed in tables, the average energy being 0.3 to 0.4 of this maximum. A discussion of this, of the loss of energy by electrons in traversing matter, and of the distribution of ions along the tracks of β particles is given in the chapter by Gray.
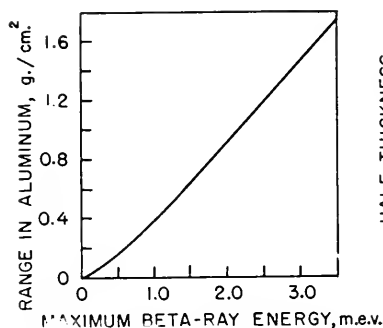
Fig. 1. Absorption of β rays in aluminum—energy of β rays vs. range in aluminum.
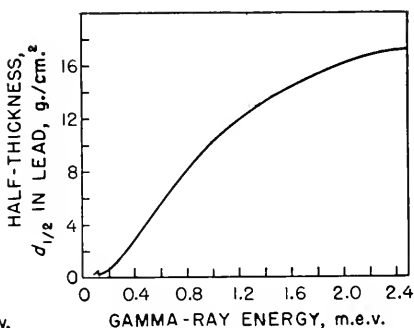
Fig. 2. Absorption of γ rays in lead—energy of γ rays vs. half-thickness in lead.

β particles from a given radioisotope are considered to have a range, but this is much less well defined than for α particles. The absorption curve for β rays (the fraction passing through a given thickness plotted against that thickness) is roughly exponential but falls off more rapidly than exponentially as the absorber thickness is increased. This range is usually listed in terms of surface density or grams per square centimeter (thickness times density), since in such units it is nearly independent of the nature of the absorber. Figure 1

is a curve of β-ray energy (maximum) versus the range of the β rays in aluminum (16).

γ rays are photons with energies that range from 0.01 to 2 m.e.v. for the important tracer nuclides. They have greater penetration than the usual X rays though the newer high potential X-ray instruments can duplicate and even exceed these energies. Their penetration is very much greater than that of the β rays and consequently their effects are not localized near the source but appear throughout the entire system. γ rays produce effects by means of the secondary electrons to which they give part of their energy, principally by means of the Compton process, or collision between a γ ray and an electron, in which the electron takes part of the energy of the γ ray as kinetic energy and the γ ray is scattered with the remainder. The absorption of γ rays in matter is not characterized by a range but is exponential and can be considered in terms of the absorption coefficient, defined as μ in the equation:

$$I = I_0 e^{-\mu d}$$

in which I_0 is the incident intensity and I the intensity after passing through a thickness d. The half-thickness, the layer necessary to cut the intensity to half its original value, is given by the relation:

$$d_{1/2} = 0.693/\mu$$

The thickness, d, is usually multiplied by the density and expressed in grams per square centimeter and μ is similarly given in square centimeters per gram. Figure 2 shows the relation between the absorption half-thickness in lead and the energy of the γ ray (16).

γ rays, in particular those of low energy, frequently are absorbed in the atom in which they originate and their energy is transferred to an electron as kinetic energy. This process, known as "internal conversion," results in an electron with energy less than that of the original γ ray by an amount equal to the binding energy with which the electron was held in its place in the atom. The electron is accompanied by X rays given off when other electrons drop in to fill its place and usually by some fraction of the original γ rays that escapes conversion.

Other processes by which radioactive atoms decay are positron (or positive β-ray) emission and K electron capture. These processes occur in nuclides with fewer neutrons than their stable isotopes and correspond to the hypothetical conversion of a proton into a neutron.

The atomic number is thereby reduced by one in contrast to the negative β-ray process, in which the atomic number increases by one.   In positron emission the nucleus gives off a positron, which has an energy distribution like that of the electron.   A positron does not have a long life; as soon as it is slowed down by collisions to nearly zero kinetic energy, it interacts with an electron in some atom to annihilate the pair and produce two γ rays, each of which has an energy equivalent to the mass of the electron at rest (0.51 m.e.v.).   If a nucleus is unstable to this type of transition (if it must reduce its charge in its decay process) but does not have enough energy to create a positron, it may capture one of the innermost (K) electrons around it.   This process is accompanied by X rays given off when the K shell is refilled, and usually by γ radiation.

An additional type of decay is found in some of the cases of nuclear isomerism (the existence of nuclear species with the same atomic number and mass but different nuclear properties), in which transition between isomers occurs.   Since no charged particle can be emitted from the nucleus without changing Z, such decay is characterized only by γ-ray emission (but, in addition, electrons may be expelled from the atom owing to the internal conversion of these γ rays, and then X rays would accompany the conversion process).

2. Radioactive Decay Law

Radioactive decay is governed by the familiar first order rate equation:

$$-dN/dt = \lambda N$$

in which $-dN/dt$ is the number of disintegrations per unit time, λ is the decay constant, or decay probability, and N is the number of radioactive atoms present.   Integrated forms of this equation are:

$$N = N_0 e^{-\lambda t} \qquad \text{and} \qquad \log_{10}(N_0/N) = (\lambda t/2.303)$$

in which N_0 is the number of atoms at time $t = 0$.   From this law it follows that the fraction decaying in time dt is λ and is thus independent of the amount present.   The rate of decay is usually stated in terms of the time required for one-half to decay; this time is called the half-life and is related to the decay constant by the equation:

$$t_{1/2} = 0.693/\lambda$$

Measured half-lives of radioisotopes range from 10^{-6} second to 10^{12} years. It is obvious that many isotopes will have decay periods much too short to allow their use as tracers, while for others the periods are so long that the radioactivity will be of too low intensity to measure. The useful range of half-lives is nearly spanned by the carbon isotopes, C^{11}, with half-life of 20 minutes, and C^{14}, 5000 years.

3. Units

The most commonly used unit for comparing quantity of radioactivity is the curie (17). The unit was originally set up so that other radioactive elements, including the daughter products of radium, could be compared to radium, for many years the reference substance. It is now widely used to mean the amount of radioactive substance giving the same number of disintegrations as one gram of radium, namely, 3.7×10^{10} per second. It is hoped that, although the experimental value of the decay rate of radium may change, this numerical definition of the curie will be retained. For most work the fractional units, the millicurie (mc. $= 3.7 \times 10^7$ disintegrations per second, 2.2×10^9 disintegrations per minute) and microcurie (1 $\mu c. = 2.2 \times 10^6$ disintegrations per minute) are closer to the range of activity being handled.

In some instances the term "milligrams radium equivalent" is used. One should check each author's definition of this term, since some use it as a synonym for millicurie while others make the comparison in this case on the basis of γ-ray activity as measured by ionization chamber (see below). The two methods may give completely different results.

To avoid the conflicting definitions of the curie, a new unit, the rutherford (rd.), has been proposed (18). It is defined on a purely numerical basis as that quantity of any radioactive material decaying at the rate of 10^6 disintegrations per second. The new unit is thus 1/37 mc. according to the usual definition of the latter. It has not had wide acceptance and seems to be considered superfluous.

The differential equation for decay $(-dN/dt = \lambda N)$ relates the activity, or disintegrations per unit time, to the weight of radioactive material, since the number of atoms, N, is equal to the weight of material divided by the molecular weight and multiplied by Avogadro's number. From these various relations such problems as the weight of isotope representing a millicurie of substance of any given half-life can easily be solved.

C. AVAILABILITY OF ISOTOPES

Radioisotopes can be produced by a number of different nuclear reactions in which one of a few bombarding particles produces a change in the nucleus of the atom, yielding either a new isotope of the same element or of a different element. The former process is termed activation and the latter is true transmutation of elements. Included in the projectiles are the neutron, proton, deuteron, and α particle and in the nuclear reaction one of these particles or a γ ray will be emitted. A projectile that is positively charged must have high energy to penetrate the positive nucleus of the atom and produce a reaction, but the uncharged neutron gives better yields if its energy is small (*1*, p. 16; *2*, p. 43; etc.).

1. Radioisotopes Producible by Slow Neutron Activation

The neutron is undoubtedly the most effective of the various bombarding particles and the most prolific reaction is that of neutron capture, in which energy in the form of γ radiation is all that is emitted. Since the existence of the neutron is highly transitory it must be produced near the material to be activated. In most instances isotopes producible by neutron activation will be available from the Atomic Energy Commission (*19*) although it is possible that some isotope of short half-life may need to be made on the spot. If this is necessary someone with proper experience will be available or should be consulted before such production is attempted.

Although some of the early work employing artificial radioactivity was done with tracers produced by neutrons from a radium–beryllium or similar source, the amounts of activities so producible are quite inadequate for biological application. A vastly higher neutron intensity can be obtained from a cyclotron by using it to accelerate deuterons and bombarding beryllium with these. Slowing the neutrons down from their original high energies to the energy of atoms and molecules at room temperature increases their effectiveness in inducing the capture (n,γ) (neutron in, γ ray out) process. Such "thermal" neutrons are produced by allowing the original high energy neutrons to traverse a few inches of paraffin "moderator" in order to reduce their kinetic energy by collision. This neutron activation reaction produces nuclides isotopic with the original element and the active atoms are thus diluted by a large number of inactive ones. In other words the "specific activity," in disintegrations per unit time

per unit weight or millicuries per unit weight, will be quite low.   One way of getting around this is to use the so-called Szilard-Chalmers process.   In nuclear processes such as radioactive decay and the emission of γ rays after neutron capture, the atom itself recoils and chemical bonds holding it in some molecule are broken.   If these bonds are covalent or of other type not permitting exchange with free atoms, it may be possible to separate the atoms that have been activated and have recoiled from the bulk of inactive ones.   The classic example of this is the production of radioactive bromine by irradiation of ethyl bromide with neutrons.   The bromine atoms recoiling after activation break the C-Br bonds, and the liberated atoms do not re-enter the organic molecules (20).

2. Other Important Radioisotopes

Except for short-lived isotopes all those producible by the neutron capture process are available from the Atomic Energy Commission (19; 2, p. 56; 3, p. 43).   In spite of the fact that the neutron intensity in the neutron chain reactor or pile is much greater than that available from cyclotrons the specific activities of some radioisotopes from the pile are not sufficient for some puposes.   This is particularly true if the parent isotope is present to only a small percentage.   For example the 44 day Fe^{59} is produced by neutron capture in Fe^{58}, the abundance of which is only 0.33% and the important 180 day Ca^{45} comes from the 2.1% abundant Ca^{44}.   Unfortunately the Szilard-Chalmers process does not work in a pile since the radiation intensity is so great that the covalent compounds necessary for the process are destroyed.   The radioactive atoms are thus diluted by many others from the decomposed molecules and the expected enrichment is not obtained.

Certain other isotopes that can be made by the n,p reaction are also available (19).   Among these are C^{14}, P^{32}, and S^{35}.   In this case the target and product are not isotopic and production methods have been devised for some of these to yield high activities undiluted by inactive isotopes, i.e., "carrier-free."   If this were actually attained the specific activity would be that of the pure isotope, but usually there will be some dilution by inactive isotopes occurring as impurities.   The products of the fission of uranium are, of course, highly radioactive and some of them are quite useful as tracers, though they are not of greatest interest biologically.   In this case also the produc-

TABLE I
Radioisotopes of Biological Interest

Symbol	Half-life[a]	Radiation and energy[b]	Methods of production[c]	Source[d]	Amount available, mc.	Sp. act. mc. g. element	Remarks
H^3	11 y	β^- 0.018, no γ	Li (n,H^3)	Pile	10–1000	High	Short half-life
C^{11}	20 m	β^+ 1.0	B (d,n)	Cyclotron	…	High	0.1 to 4% of C^{14}
C^{14}	5000 y	β^- 0.156, no γ	N (n,p)	Pile	1	0.5–20	Short half-life
N^{13}	10 m	β^+ 1.24	Several	Cyclotron	…	…	
Na^{22}	2.6 y	β^+ 0.6	Mg (d,α)	Cyclotron	…	…	
Na^{24}	15 h	β^- 1.4, γ 1.4, 2.8	Na (n,γ)	Pile	20	150	
P^{32}	14.3 d	β^- 1.7, no γ	P (n,γ); S (n,p)	Pile; Pile	350; Large	50; C.F.	Available as bombarded or separated from S
S^{35}	87 d	β^- 0.17, no γ	S (n,γ); Cl (n,p)	Pile; ……	Large; 1	4×10^4; C.F.	As sulfate; As sulfide
Cl^{36}	10^6 y	β^- 0.64, β^+	Cl (n,γ)	Pile	0.005	5×10^{-4}	Long half-life, low activity
Cl^{38}	38 m	β^- several	Several	Cyclotron	…	…	Short half-life
K^{42}	12 h	β^- 3.6, γ 1.5	K (n,γ)	Pile	130	10	
Ca^{45}	180 d	β^- 0.26, no γ	Ca (n,γ); Sc (n,p)	Pile; Pile	0.8; 1	0.1; C.F.	Not in routine production
Cr^{51}	26 d	e^-, γ 0.32	Cr (n,γ)	Pile	50	50	
Mn^{52}	5.8 d	β^+ 0.8, γ 1.0	Cr $(d,2n)$	Cyclotron			
Mn^{54}	310 d	e^-, γ 0.8	Cr (d,n)	Cyclotron			
Fe^{55}	4 yr	e^-, no γ	Fe (n,γ)	Pile	0.9	0.05	
Fe^{59}	46 d	β^- 0.26, γ 1.1	Fe (n,γ)	Pile	1	0.06	
Co^{60}	5.3 y	β^- 0.3, γ 1.1, 1.3	Co (n,γ)	Pile	20	30	
Cu^{64}	13 h	β^- 0.6, β^+ 0.7	Cu (n,γ)	Pile	100	300	
Zn^{65}	250 d	β^+ 0.3, γ 1.1	Zn (n,γ)	Pile	15	1.9	

Symbol	Half-life[a]	Radiation and energy[b]	Methods of production[c]	Source[d]	Amount available, mc.	Sp. act., mc./g. element	Remarks
As^{74}	17 d	β^- 1.3, β^+ 0.9, γ 0.6	Ge (d,n)	Cyclotron	25	560	
As^{76}	27 h	β^-'s 1 to 3 γ's	As (n,γ)	Pile		C.F.	
As^{77}	40 h	β^- 0.8, no γ	Ge (n,γ)	Pile	0.7	C.F.	Produced in Ge, can be separated
Se^{75}	127 d	e^-, γ 0.5	Se (n,γ)	Pile	65	3.3	
Br^{82}	34 hr	β^- 0.46, γ's	Br (n,γ)	Pile	70	120	
Sr^{89}	53 d	β^- 1.5, no γ	Sr (n,γ)	Pile	1.5	0.11	Also available as F.P. Also contains Sr^{90} and Y^{90}
Sr^{90}	30 y	β^- 0.65, no γ	F.P.	Pile	…	C.F.	With Sr^{89} and containing Y^{90}
Y^{90}	62 h	β^- 2.2, no γ	F.P.	Pile	1	C.F.	Daughter of Sr^{90}
Ag^{110}	225 d	β^- 0.6, γ's	Ag (n,γ)	Pile	35	C.F.	
Ag^{111}	7.5 d	β^- 1.0, no γ	Pd (d,p)	Pile	10	8	Daughter of Pd^{111}
I^{130}	13 h	β^- 1, γ 0.7	Te $(d,2n)$	Cyclotron		C.F.	
I^{131}	8 d	β^- 0.6, γ 0.4, 0.08	Te (n,γ)	Pile	130	C.F.	Mixed with Te parent, and Te activities or separated
Ba^{140}	13 d	β^- 1.05, γ 0.5	F.P.	Pile	1	C.F.	With La daughter
La^{140}	40 h	β^- s, γ's	F.P.	Pile	1	C.F.	Can be separated from Ba^{140}
Au^{198}	2.7 d	β^- 1.0, γ 0.4	Au (n,γ)	Pile	80	5×10^3	

[a] m = minutes.

[b] It should be remembered that all positron emitters are also sources of annihilation γ radiation of 0.5 m.e.v. energy. Energy values for β radiation are maxima for various spectra.

[c] F.P. means isotope is a fission product and is separated from uranium and other fission products before shipment.

[d] All pile producible isotopes can be obtained from the Isotopes Division, the amounts readily obtainable being listed in column six and the specific activity in column seven (C.F. denotes carrier-free). These two columns should not be considered the final word; greater intensities and higher specific activities can be obtained in some cases. For further clarification see (19). For additional information on these and other isotopes see (21,22).

tion method does not put a limit on their specific activity. See Table I for information on the biologically important radioisotopes.

There are a number of important isotopes that cannot be produced by neutron bombardment, but must be made in a cyclotron (*3*, p. 52). Some of these are included in Table I. Those producible by cyclotron have not been applied extensively to tracer research, but they doubtless would be if they were readily available. Many of them are not limited in specific activity, if chemical separation can be devised. For further detail consult the well known table of isotopes by Seaborg (*21,22*).

D. DETECTION AND MEASUREMENT OF RADIOACTIVITY

The use of radioactive tracers is dependent upon the measurement of their radioactivity, and the accuracy obtainable by using them is only as good as the accuracy of this measurement. Fortunately this is high in comparison to the other variables in biological research and the method is, in general, capable of results as accurate as the research requires (*2*, pp. 105, 141).

Radioactivity is measured usually by the ionization produced by the radiation. A large variety of instruments has been applied to the problem and is discussed fairly extensively in the literature. Only a few of these will be considered here, first as applied to the general problem of radiation measurement and, second, to the special problem presented by isotopes whose radiation is of very low energy, particularly C^{14}. Three general principles of measurement are used: photography, ionization chambers, and counters. The photographic method does not have high sensitivity, but often it does give the answer when no other method will. Its present use is largely confined to two fields: radiation monitoring and radioautographs. Both of these are considered in later sections.

1. Ionization Chambers

An ionization chamber is a gas-filled space in which two oppositely charged electrodes collect the ions produced in the gas by radiation. This collection of ions by what is essentially a charged condenser constitutes a small electric current measured either as a current or as the rate at which the electrodes are discharged. A number of conventional circuits using various types of electrometers have been widely used. With the development of electronics it has become possible to

amplify the minute currents produced by small amounts of radioactive substances (23) and many instruments employing this method and a variety of circuits are now on the market.

The most important application of this method is in monitoring instruments, discussed in Section G of this chapter. A few more accurate instruments using this principle are available (82) and the circuits of others have been published. Instruments of this type generally give a reading on a meter which is directly a measure of the rate at which the radiation is entering the chamber, hence they are called "rate-meters." Servicing the circuits in these instruments is not a job for an amateur, but most of the manufacturers will make necessary repairs on instruments and the newer models are becoming more reliable.

2. Electroscopes

Though the older metallic leaf electroscopes can still find some application, they will be found too insensitive for many research problems. Their use has largely been supplanted by the quartz fiber type of which the Lauritsen electroscope (25) is the most widely known example (83). This is an extremely useful instrument, quite moderately priced, relatively rugged, and very easy to use. Since it is felt that this is the best all-round instrument for a limited research budget, it will be discussed at considerable length.

The principle of an electroscope of either metal leaf or quartz fiber type is that of an electrical condenser in which the case is one conductor and the leaf or fiber system the other. An insulator of sulfur, amber, or a plastic of high resistivity supports the moving system within the case. The quartz fiber of the Lauritsen is made conducting by a sputtered metal coat and is connected to and (when discharged) is parallel to a metal wire. An unstable condition is produced in the fiber system and when it is charged the fiber will be bent away from the wire; as the charge is neutralized by the collection of ions, the fiber regains its normal position (see Fig. 3).

Certain other features of the Lauritsen are important. There is a T on the end of the fiber, which is brought into focus on the scale of the optical system, and the rate at which this T crosses the scale is timed by a stop watch and recorded. A charging button outside the case makes contact between the fiber system and the external charging circuit that provides the necessary voltage. Though elec-

trostatic chargers are available, a small power pack is much more reliable for providing this potential.

The scale of the electroscope is not linear throughout its length so that it should be calibrated, or only one part of it, preferably the first third, should be used. In the absence of any radioactive substance, an electroscope fiber will still discharge at a slow background rate. The magnitude of this background will depend on conditions but is usually 0.05 to 0.25 scale divisions per minute. It is caused by cosmic rays, minute amounts of radioactive substances near and in the instruments, and by leakage across the insulator. After the production of ion pairs by radiation, there is always a possibility that the ions will recombine and not be measured. This possibility increases
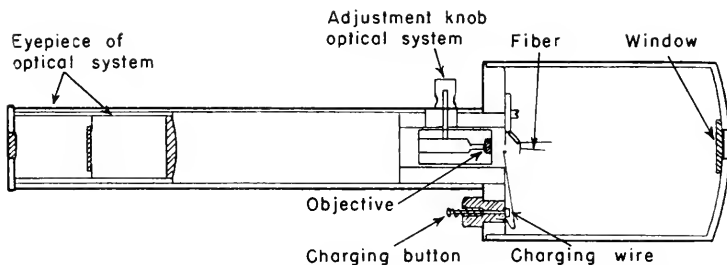
Fig. 3. Lauritsen electroscope. (Courtesy *F. C. Henson Co.*)

with increasing density of ionization, so that as the intensity of radiation increases the electroscope no longer collects as large a fraction of the ions. There is then a limit above which the response of the electroscope will not be proportional to the amount of radioactive substance. One should either stay below this point or make a correction. For a particular case the point at which an error greater than 1% was introduced by this nonlinearity was found to be 60–70 scale divisions per minute.

Numerous modifications of the chamber of the Lauritsen electroscope have been made for different purposes (2, p. 119). The instruments come with a cylindrical chamber of sheet aluminum or a similar shell with a thin aluminum window, to permit the measurement of low energy β rays. Other modifications available from the manufacturer include a brass chamber into which the sample can be put.

A chamber, preferable to those on market, can readily be made in the shop. Its advantages are rigidity, more favorable geometrical

arrangement, and a drying agent to improve the performance in humid weather. The latter can cause serious interference since the background in a humid atmosphere can vary as much as tenfold. The chamber to be described can be machined in any instrument shop from a 3–3¼ in. cube of aluminum (see Fig. 4). A 2–2¼ in. square hole is machined through this cube in what will be its vertical direction. This produces a square shell with about ½ in. walls that can
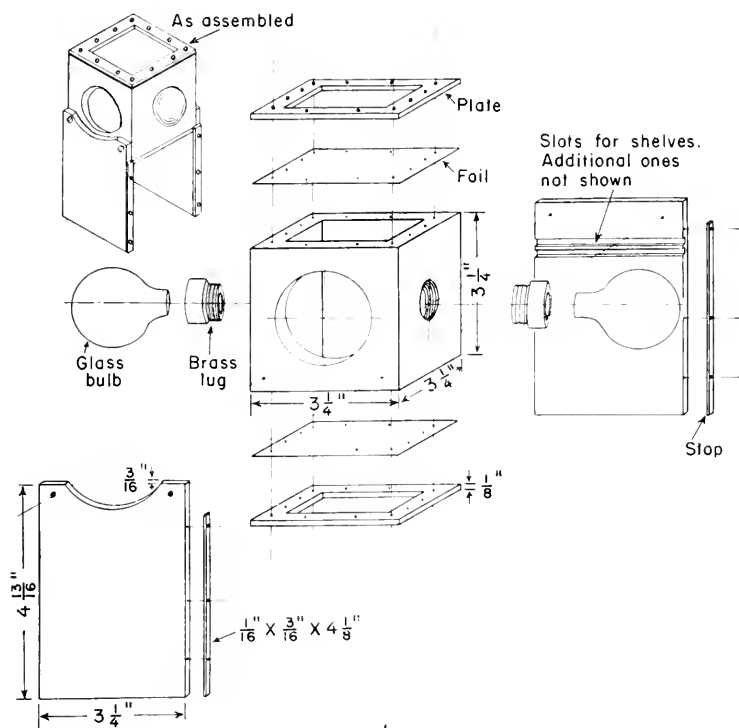
Fig. 4. Cubical cast aluminum ionization chamber for Lauritsen electroscope.

also be cast directly if casting facilities are available. At right angles to this hole are machined a circular hole with the diameter of the electroscope head so that a snug fit is obtained over this head, and opposite to this a ½ in. hole over which a piece of etched glass is cemented as a window for the light source. Through the remaining two walls two 1 in. holes are drilled and tapped with fairly large threads. Lugs to fit these holes are made of brass or other metal tubing. In these

lugs are cemented glass bulbs that can be filled with a drying agent. By means of rectangular retaining rims, made of $\frac{1}{8}$ in. sheet aluminum, aluminum foil windows of any desired thickness can be held securely on the top and bottom of the case (see Figs. 4 and 5). The system can be made fairly impervious to air by applying a soft wax
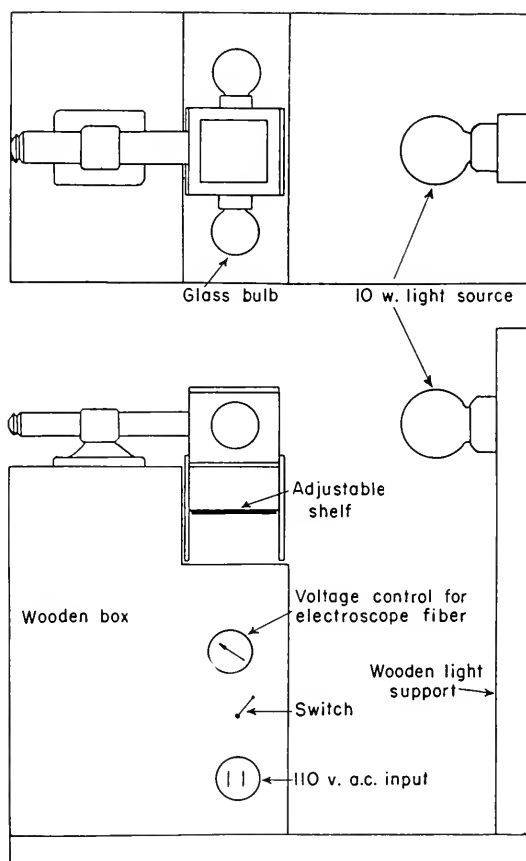
Fig. 5. Lauritsen electroscope assembly.

(e.g., Apiezon) at the places where joints are made and cementing the aluminum windows in place with Duco cement.

As support for this chamber, two vertical aluminum plates are made from $\frac{1}{8} \times 3$ in. strips long enough to reach the level of the table. Grooves machined in these plates allow an adjustable shelf to be

placed at various heights thus allowing a change in the solid angle and providing a much wider range of usefulness. These supports can be fastened to the chamber with screws that fit into drilled and tapped holes in the chamber. A wooden box of proper height to bring the optical system to eye level can also contain the charging power pack, and a voltage control and switches can be located conveniently on this box.

A number of other types of quartz fiber electroscopes are now on the market. In the main these are designed as monitoring instruments and are discussed in Section G of this chapter.

3. Counters

As the voltage on a gas-filled chamber is increased, the chamber undergoes several distinct types of behavior toward radiation. A similar effect is observed if the gas pressure in the chamber is gradually reduced. These changes bring one to the region of counters, in which individual ionizing events due to single particles can be counted (26). In obtaining the desired characteristics in the counter, the separation and form of the electrodes and the composition of the gas as well as the pressure and voltage are of importance. At voltages around 1 kv. with air or nitrogen at one atmosphere and with electrodes that are parallel plates separated by 0.5 to 1 cm., an individual α particle will produce enough ionization so that it can be counted. When the particle passes through the air gap it produces electrons and positive ions, which travel to the electrodes rapidly and rapidly reduce the voltage across the gap. This reduction of voltage constitutes a pulse, which is terminated after the ions are collected. The system then recharges itself and is ready for the next α particle. Because of the much lower ionization that they produce within this short path length, β particles do not give enough of a pulse and only α particles can be counted. Even so, the chamber requires high amplification, which can be given by a linear amplifier circuit.

Increasing the voltage on the system brings a region in which the electrons, the negative half of the ion pairs produced by the incident radiation, are caused by the electrical field to move so rapidly that they produce additional ion pairs. This process of gas amplification (referred to as an avalanche) is aided if the positive electrode is a straight wire and the negative electrode a cylinder surrounding it. With the aid of gas amplification the further electronic amplification

does not need to be as great and the individual pulses produced by radiations of all kinds can be counted. There is a voltage region in which the size of the pulse given by the counter is proportional to the original amount of ionization and hence is different for the different types of radiation. This so-called proportional region is useful for counting α-active material in cases in which it is desirable to count α particles above a high background of β radiation. A satisfactory arrangement (84) uses methane at atmospheric pressure as counter gas, a wire down the center of a cylinder of 1–2 cm. diameter as counter with a potential of 2500 v. The samples are placed directly in the counter. The circuit used with it contains a pulse height selector along with the necessary amplifier (27).

An important use of this principle is in neutron counters. The element boron is highly sensitive to neutrons giving instantaneously the nuclear reaction:

$$_5B^{10} + _0n^1 \longrightarrow {_2He^4} + {_3Li^7}$$

A considerable energy is also given off as kinetic energy of the product particles. This results in a large amount of ionization, producing a pulse. This instrument is capable of measuring neutrons in the presence of high intensities of other radiation, since no other radiation will yield as much ionization per particle. The boron may be in the form of a coat of the free element on the inside wall or of a filling of BF_3 gas (85).

Geiger-Mueller Counter. A further increase in the voltage or, better, a reduction of the gas pressure to about 10 mm. Hg, at which pressure the required voltage is 1000–1500, brings the tube to the Geiger counter region. Although more properly termed a Geiger-Mueller counter tube, current usage favors the shorter term "Geiger counter" or the abbreviation "GM tube." The gas amplification becomes greater, and the avalanche spreads completely along the wire. This ionization is so extensive that the tube will continue to discharge if the voltage on the counter electrodes is maintained at its original value. However, when the discharge occurs, the voltage across the counter drops below the discharge point, and the avalanche stops; when the voltage recovers its original value, the tube is ready for another particle.

This sudden lowering of the potential constitutes the pulse. It is taken off the positive center wire of the counter and hence it is negative in sign. The size of the pulse is determined by the amplification

process inside the counter and is independent of the number of ions produced by the original particle. Thus no discrimination can be made among the various types of radiation by the size of the pulse (28). If one plots a curve of the observed counting rate for a given source *vs.* the voltage, the curve has the characteristic shape shown in Figure 6. The level portion, known as a plateau, is the operating region for the Geiger counters. The length and slope of the plateau are important characteristics of the tube.

Geiger counters are usually divided into two types, depending on the nature of the process by which the discharge is stopped. If the
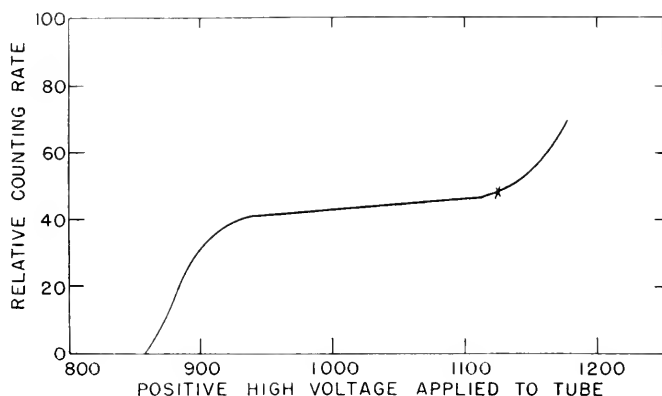
Fig. 6. Counting rate *vs.* voltage—"plateau" curve for Geiger counter.

gas contains about 10% of an organic vapor such as ethyl alcohol, diethyl ether, an ester, or a hydrocarbon along with 90% of a rare gas (argon or neon), the stopping of the discharge is aided by intermediate steps that lead to disruption of the organic molecules. Such a counter is said to be self-quenching. If the gas does not contain an organic vapor no such mechanism will help and an external electronic circuit is needed to lower the potential across the counter rapidly after the discharge starts. Such a circuit is the typical Neher-Harper circuit shown in Figure 7*b* (29).

In the firing of a Geiger counter the potential drops as a result of the current carried by the original and secondary ions across the high voltage. This current causes the potential to drop, either unassisted in the conventional resistance–capacitance coupling of a self-quenching counter (Fig. 7*a*) or with the help of a Neher-Harper

type of circuit.   When the potential has fallen sufficiently to stop
the pulse, the counter voltage again is built up to its maximum and a
second pulse can be received.   It can be seen that there would be a
time during which the counter is not able to receive pulses.   The
"dead time" is quite important in that, as the number of counts per
unit time is increased, the spacing between them decreases and a
larger fraction will occur within spacings too short for the counter
to have recovered.   These counts will be lost and their loss should be
corrected for.   There is a straightforward way of determining the
dead time and using it in making this correction (40).   An empirical
correction that has found considerable application for counting rates
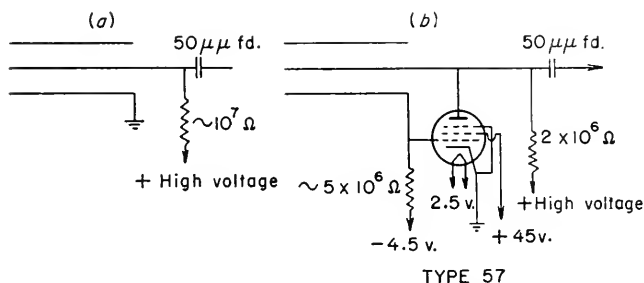
Fig. 7.   Geiger counter quenching circuits.   (a) Resistance–capacitance
coupling.   (b) Neher-Harper quenching circuit.

under 5000 per minute is 1% per thousand (1% at 1000, 2% at 2000,
etc.).

Mention has been made of the type of gas used to fill counters.
Almost all the commerically available GM tubes are filled with a
self-quenching mixture of about 9 cm. mercury pressure of argon
or other rare gas and 1 cm. of an organic vapor (ethyl alcohol, diethyl
ether, or an acetate ester).   The use of a Neher-Harper circuit is,
in general, unnecessary, though counting characteristics may be
improved.   The disadvantage of the self-quenching type of tube is
that the quenching mechanism breaks down the organic molecules.
Since a limited number of such molecules can be put into a tube, there
will be a limit to tube life, usually given as 10^8 to 10^9 total counts.
A non-self-quenching tube without such a complex vapor does not
have this limited life.

Several different types of construction are used in GM tubes.
One of the sturdiest is the all-glass tube in which the cathode is a

silver mirror deposit on the inside of the cylindrical glass jacket and
the anode is a tungsten wire stretched down the center.   From the
name of an early manufacturer it is known as an Eck and Krebs tube
(see Fig. 8).   The minimum wall thickness that can be used in this
type of construction is about 0.15 mm., which will completely
stop β radiation of energies less than 0.17 m.e.v. and will absorb
a large fraction of the β rays unless their energy is at least 0.4 m.e.v.
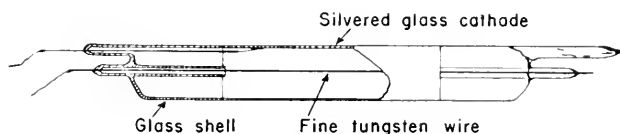
Fig. 8.   Cylindrical glass GM tube.
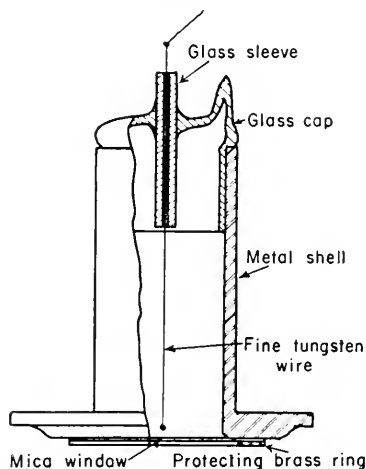
Fig. 9.   Mica end window counter of bell type GM tube.

There are a number of commerical sources for this type of tube
(85–87) and, since there is much less that can go wrong with this kind
of construction, the tubes are recommended for work with energetic
β emitters.   Holders with adjustable shelves can be obtained (85)
and thick-walled tubes of the same sort that will allow only γ rays
and the most energetic β rays to pass are also available.

For work with emitters of low to average energy β rays, the mica
end window counter is generally used (see Fig. 9).   In this type the
cathode is a metal cylinder that may be the shell of the tube or may be

inside a glass shell. The center wire, usually tungsten, is supported on one end only and has a small glass bead on the free end to prevent sparking. The window affixed to a metal or glass flange on the open end of the cylindrical tube is of high grade mica split to the desired thickness, which can be anything from 10 μ up. The minimum allowable thickness is that which will support a vacuum and will depend on the area of the window. The window must be of uniform thickness and it will be if split along cleavage planes. If not uniform, lines marking a sharp variation in thickness will be visible.

End window counters are available from a number of commercial sources (*84–87*), and additional manufacturers are coming into the market at the present time. So many different types of construction are available that various recent catalogs should be consulted. These tubes are expensive, particularly so since their life is limited and the windows are fragile. Refilling services are offered by the manufacturers, but there may be some question as to the value of refilling a tube with fresh gas after the quenching gas has been decomposed, since other effects such as pitting of the anode may also occur. Tubes similar to these can be machined and built in the laboratory, but there is a considerable art to the making of good tubes, which needs to be acquired before any consistent success can be expected. A description of such a tube is given by Copp and Greenberg (*30*). It is not expected that the small laboratory entering this field will have the time or patience to develop the art. Although the commerical cost may seem high, considerable experience is necessary before a reasonable percentage of homemade tubes will operate successfully. (For further discussion of Geiger tube construction and action see among others the following references: *1*, p. 58; *2*, pp. 105, 141; *28,31*.)

Counter Circuits and Scalers. Some way of recording the counts must be available. The pulses from a Geiger counter come in much too rapidly and are too randomly spaced to be recorded efficiently by a mechanical counter, however rapid its action. For this purpose an electronic circuit known as a "scaler" is always used. Such a circuit is composed of a number of units, each of which will reduce or scale down the number of pulses reaching it by a factor of two. A "scale of two" will record the first pulse reaching it by lighting a neon glow lamp. In receiving the second pulse the glow lamp is extinguished and a pulse is sent on to the next scaling stage. The fraction of reduction in counts will be equal to 2 raised to a power, that power being the number of scales of two in the circuit. The most commonly used circuits employ six such scaling stages and are,

thereby, scales of 64 (i.e., 2^6). The pulses up to 64 are recorded on the neon lights and the sixty-fourth count clears the lights and is recorded on the mechanical register. Decade scaling circuits are also available, but have not been used as widely (90,91).

Most of the scaling circuits commercially made for use in this field include, in addition, a regulated high voltage supply for the GM tube. Regulation is essential since the counting rate varies with changing voltage even in the region of the plateau. High counting rates cause a drain on the high voltage supply, which may cause difficulty if the latter is not well regulated against this. A particularly annoying source of trouble is an unsteady house supply of electricity. During periods of overloading, the actual a.c. voltage delivered to the set may be as much as 20 v. below the rated value. If this situation exists some form of stabilizer should be added, since the circuit cannot be expected to operate satisfactorily under such conditions. Electronic stabilizers now on the market appear to be quite adequate for both voltage and frequency fluctuations (96). A small motor generator set will take care of line voltage changes but will introduce fluctuations in voltage if the frequency of the line voltage is not constant.

Scalers including high voltage supplies and employing the so-called Higinbotham circuit (32), which is new and yet thoroughly tested through Atomic Energy Commission use, are manufactured by several companies and are at present available, with a built-in mechanical register, for about $400 (84,88,89). If the research budget can afford these, the problem of counter circuit is nearly solved—nearly, because these are not perfect and they do require adjustment and repairs. The latter are usually not serious, but some one conversant with electronics should be available.

4. Special Measuring Techniques

In some cases of particular importance the β radiation of the radioisotope is of such low energy and penetrating power that even the thin window of the end window counter absorbs a large part of the radiation. If it is necessary to measure samples of appreciable weight, the self-absorption of the rays by the sample can be quite serious. In order to circumvent these difficulties for the three important tracers H^3, C^{14} and S^{35}, various techniques have been developed. In the case of tritium, H^3, difficulties in its production and measurement as compared to the ready availability, low cost, and easy

measurement of deuterium combine to lessen the importance of tritium as a biological tracer. Though some instruments for measuring its 0.018 m.e.v. β particle have been described, they will not be discussed here. The other two of these isotopes, C^{14} and S^{35}, are more easily made and so readily available that considerable thought has been given to efficient ways of measuring them, particularly C^{14} (33–37; 1, p. 172).

One solution is their introduction directly into the counting chamber as a gas. In the case of carbon, the obvious gas that can be obtained as the end product in any process is carbon dioxide. This is fine for ionization chamber measurements since the chemical and physical properties of carbon dioxide make it easy to handle in a gas manipulating system. One such instrument described is based on a Lauritsen electroscope and is reputedly easy to use and trustworthy (34,37). The current produced across a charged spherical quartz ionization chamber by the radioactive carbon dioxide put into it is transmitted to the working parts of a Lauritsen electroscope where it is measured. Modifications of this idea for H^3 and S^{35} measurements are also described. There are reports in the literature on a number of other ionization chambers into which $C^{14}O_2$ samples can be introduced.

A GM counter will not operate in its normal counting region if appreciable amounts of oxygen, chlorine, carbon dioxide, or any other gas containing a high percentage of an electronegative element are present, so that one cannot introduce a $C^{14}O_2$ sample into such a counter normally. Since a counter has a considerably greater sensitivity than an ionization chamber, it has been a goal to obtain a counter into which a C^{14} sample could be introduced. Although this is possible if the C^{14} is present in methyl alcohol or a similar compound, the difficulties in the preparation of the latter from the material in which the tracer is obtained from the experiment are too great. It has been found that by adding a small percentage of carbon disulfide vapor to the carbon dioxide a stable counting region can be obtained with voltages in the neighborhood of 3–4 kv. (35,36). Since the maintenance of such voltages across resistors is a serious problem and the technique of gas handling is not simple in itself, this does not appear on the surface to be a solution to the problem that would appeal to biologists.

C^{14} users are coming more to the opinion that the mica window counter tube offers the best solution to their problems (3, p. 104).

Commercial tubes with window thicknesses down to 1.5 mg. per square centimeter are available ($84,85$, etc.) at somewhat higher prices than are charged for those with thicker windows. The range of the C^{14} β rays is 25 mg. per square centimeter and following a rough rule that range is about eight times the half-thickness of β radiation at the beginning of the absorption curve, ~ 3 mg. per square centimeter of total absorber would cut the intensity to half, 6 mg. per square centimeter to one-fourth, and so on. An estimation of the loss to be expected can be made on this basis.

A rather expensive screen wall counter based on a design by Libby (33) is commercially available (85). Samples are introduced into the vacuum chamber and the cathode is a cylindrical screen between the sample and the center wire. Since only one sample can be measured at a filling, its use is quite cumbersome. The design of counters that can measure the tracers while they are in solution has been undertaken by a number of workers and is common practice in some laboratories. The case for dipping counters is discussed by Bale (38) and by Solomon and Estes (39).

E. ERRORS AND CORRECTIONS IN MEASUREMENTS OF RADIOACTIVITY

Interpretation of the data from measurements of radioactivity requires a knowledge of the possible errors, the corrections that need to be made, and the limits of accuracy of which the method is capable (40). Since the corrections need not be made to a degree greater than the inherent accuracy, a discussion of this point will be given first.

1. Statistics of Counting

It is important in the interpretation of the data to be able to decide if a result is statistically significant. Although the subject of statistics is one of considerable complexity and is beyond the range of this chapter, a brief discussion of this point follows (41).

The disintegration of a large number of the same kind of radioactive atoms is randomly spaced in time. The counts observed from a source containing such atoms will have a random distribution obeying the Poisson distribution law. To avoid some of the difficulties in applying Poisson distribution, the normal or Gaussian distribution

law is frequently assumed to apply. If the number of counts is more than a very few, the two give essentially the same results.

The reliability of a measurement of the strength of a radioactive source depends on the number of counts observed. The function involved is the square root of the number of counts, which is multiplied by a factor that depends upon the accuracy limits one wishes to set up. According to the distribution laws, a definite probability, P, exists that the true value of a measured quantity, such as the number of counts recorded by a counter, is within the limits stated as the limits of error of the quantity. The probability that the true value is outside these limits is obviously $1 - P$. In order that the limits express a particular probability a factor, K, is chosen by which the square root of the number of counts is multiplied, the result being expressed as $C \pm K \sqrt{C}$, where C is the number of counts recorded in time t. For the standard deviation, $K = 1$ and $P = 0.683$, that is, there is a probability of 0.683 that the true value will be within the region $C \pm \sqrt{C}$ and of $1 - 0.683$ ($= 0.317$) that it will be outside this region. For the "probable error" P is chosen as 0.5 and K is 0.6745. If 90% certainty is desired ($P = 0.90$) K is equal to 1.64 and for 99% certainty ($P = 0.99$) $K = 2.58$. In the following discussion the standard deviation will be used but any other degree of certainty could be substituted, by inserting the factor K before the square root sign.

Since the standard deviation in the number of counts, C, is $\sqrt{C}$, that of the counting rate, C/t, will be $\sqrt{C}/t$ and the fractional standard deviation in either will be $\sqrt{C}/C$ or $1/\sqrt{C}$. Table II shows the per cent standard deviation for various total counts observed.

TABLE II
Standard Deviation

Total counts observed	10	40	100	400	10^3	5×10^3	10^4	4×10^4	10^5	10^6
Standard deviation, %	32	16	10	5	3.2	1.4	1.0		0.5	0.31	0.10

If more than one measurement involving some uncertainty is involved, the statistics rules for determining the uncertainty in the final answer should be applied. For example, each counter has a background in the absence of a source which must be subtracted from the measured counting rate to obtain the true counting rate due to the source. If the number of counts recorded in taking the

background measurement is B and the duration of this measurement is t', the background counting rate and its standard deviation will be $(B \pm \sqrt{B})/t'$. The true counting rate of the source, R, and its standard deviation will be given by the expression:

$$R = \frac{C}{t} - \frac{B}{t'} \pm \left(\frac{C}{t^2} + \frac{B}{t'^2} \right)^{1/2}$$

or, in words, the resultant standard deviation will be equal to the square root of the sum of the squares of the standard deviations of the measured and background rates.

It can be seen from the above that the error may be larger if the measured activity is similar to the background than if it is considerably greater. For this reason and to reduce the time required for a measurement it is desirable to deal with activities at least of the order of ten times background. For lower activities it may be well to consider the length of time that should be alloted to the measurements of the activity and of the background. If counter time is limited these should be adjusted so that each measurement has a similar error. In general, however, counter time is not limited to such an extent since the background measurement may be taken overnight when the counter is not in use for other purposes. Such a measurement is applicable to correct the daytime measurements unless the amount of active material near the counter varies from day to night.

Another point to which statistical reasoning gives emphasis is that the background should be made as low as possible, if samples of low activity are to be measured. Since a large part of this background is due to cosmic rays and to radiation from samples near the counter, shielding with lead or other dense material will reduce the background of a GM tube considerably, frequently to about half its unshielded value. Lead shields, available commerically, include mounts for the tube and samples and have only the disadvantage of expense (84,85). Storage of samples and all other radioactive materials at a distance from the counter, preferably in another room, eliminates one source of background radiation. Construction of the counter and allied equipment with materials scrupulously free of radioactivity is of utmost importance. Counting equipment should only for very special reasons be used in a room in which radiochemical operations are done, since this is a rather sure way of contaminating the counter.

The statistical approach is not, in general, applied to measurements

with the integrating type of radiation meters, such as ionization chambers and electroscopes. Since these instruments are less sensitive, the number of disintegrations required to give observable measurement is large enough so that the statistical error is less than the instrumental errors. With the Lauritsen electroscope, for example, it may be necessary to take the average of five to ten readings in order to obtain an average that seems satisfactorily to represent the mean of the individual values. These values frequently fluctuate considerably, which is probably, but not necessarily, due to statistics. If one attempts to measure the activity of an α emitter by placing a source of the proper activity directly inside a Lauritsen one finds very large fluctuations, which are undoubtedly statistical. They are larger than those from β emitters because the ionization produced by each α particle in the chamber will be very much greater (ten to fifty times) than that from a β ray so that many fewer α rays will give a similar rate of fall of the fiber.

As the intensity of radiation incident upon it increases each type of instrument reaches a point at which it no longer responds linearly but gives a reading that is too low. Discussion of the limits of the instruments and the necessary correction is given in Section D2 for the Lauritsen electroscope and in Section D3 for GM tubes. For other instruments calibration measurements should be made.

2. Errors in Sample Mounting

Other possible sources of error are in the lack of reproducibility in sample mounting, position, distribution, and thickness. The first three of these can generally be eliminated by standardizing the mounting procedure and geometrical arrangement. One point to be examined is the sample mount; samples to be compared should all be backed by the same thickness of the same material, unless there is actual experimental evidence to show that, for the particular case concerned, the backing is unimportant or can be corrected for. The necessity for this precaution is the phenomenon of back-scattering. β radiation, originally traveling in the direction away from the counter is reflected by collisions with electrons and nuclei in the mount. Such reflections send it back into the counter and increase the counting rate over its value on a thin mount or one that back-scatters less. The back-scattering of electrons (β rays) is greater as the atomic number of the elements in the mount increases. Up to a point it also

increases with increasing thickness of mount, but this effect soon reaches a maximum value. For absolute measurements it is important to avoid back-scattering completely, but since biological tracer measurements are always relative the easier alternative of constant back-scattering will suffice.

Other precautions that should be taken to avoid errors are that samples should cover the same area and that the counter arrangement should include a method for obtaining reproducible geometry. Since holders for GM tubes that include reproducible position sample mounts are now commercially available, the latter points should cause no difficulty.

3. Self-Absorption Error

One source of error that cannot always be eliminated is that due to varying sample thickness. Since the energy of the β radiation
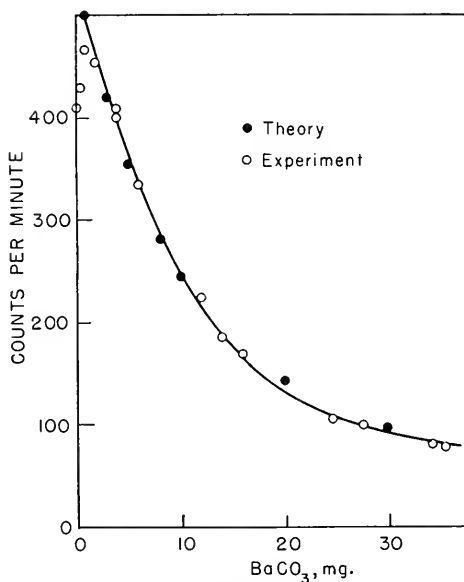
Fig. 10. Decrease in measured activity of C^{14} sample due to self-absorption (3).

from several of the most important tracer isotopes is quite low, the absorption of this radiation in the sample itself, called self-absorption, will cause trouble (33,42; 3, p. 104). With such low energies the

differences in sample thicknesses do not have to be large to have a noticeable effect on the counting rate. For example, a thickness of about 3 mg. per square centimeter would cut the observed C^{14} activity by half (see end of Sect. D4). Only half of those β rays originating from the bottom of a source of this thickness would reach the top and the net observed activity would be cut by 20–30%.

As the weight of inactive carrier present with a given amount of a radioactive substance is increased, the measured activity may increase slightly at first, then decrease fairly rapidly, and then more slowly (see Fig. 10). The original increase is due to scattering, in this case scattering of the radiation into the counter by the additional carrier atoms. The decrease that follows can be approximated by a mathematically derived equation assuming a logarithmic absorption of the radiation. Since β rays only approximately obey such an equation and particularly since this equation considers only the absorption and neglects the scattering, the correction curve obtained in this way is not completely satisfactory. The equation so developed is:

$$A = A_0 \ (1 - e^{-\mu d})/(\mu d)$$

in which μ, the absorption coefficient, is equal to $0.693/d_{1/2}$ (see Sect. B1). A_0 in this equation is the activity in the absence of carrier, while A is that activity in the presence of carrier having a thickness d. This thickness can be expressed in millimeters, milligrams per square centimeter, or other unit provided μ is expressed in the same units. A simpler correction that may be applied as an approximation is to consider that all the radiation comes from a layer midway between the top and bottom of the sample and then to apply a correction based on the thickness to which this corresponds. Thus, if the thickness of the sample is equivalent to $2\,d_{1/2}$, the ratio A/A_0 calculated on the latter hypothesis would be 0.5, while from the previous expression one would obtain $A/A_0 = 0.54$.

However, since this expression does not take into consideration the scattering it is better not to rely upon it but to determine the correction curve experimentally for the particular activity, carrier substance, and geometry involved. Such a curve as Figure 10 can readily be obtained by taking equal aliquots of a solution containing the activity, adding varying known amounts of carrier in solution, precipitating, drying, mounting, and counting in the manner chosen as standard for the experiments. The observed activity is then

plotted *vs.* sample weight (since the area is the same in all the samples, the weight and thickness are proportional) and this curve is used to obtain the corrections, using the weights of the various samples measured under these conditions.   Correction can be made to zero sample weight (thickness) or to some chosen value.

Another approach to the problem, which applies only to nuclides that are solely β emitters, is to approximate an infinitely thick source. As the weight of material of a given specific activity mounted on a constant area is increased, the measured activity will first increase and then level off approaching a constant value when the total thickness (including air and counter window) is equivalent to the range of the β particle.   For C^{14}, with a range of 25 mg. per square centimeter,
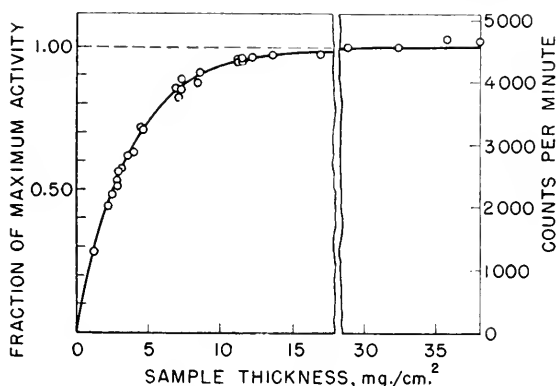
Fig. 11.   Increase in measured activity with increasing sample thickness for material of constant specific activity (42).

window thickness of 2.5 mg. per square centimeter, and 1.1 cm. (1.5 mg./cm.²) of air between source and counter, a sample of thickness of 21 mg. per square centimeter would be an "infinitely thick" source. The relative measured activity of a number of such thick sources will, it can be seen, be proportional to the *specific* activities of the samples and not to their total activities.   One technique suggested for handling such cases when the weight of active material is large is to have a number of uniform dishes or sample pans and to fill them to the top with the active material for counting.   In this way the specific activity will be determined directly in the counting operation. This will only be a relative method, but very little biological tracer research will involve absolute measurements.   If in using this method

a sample of insufficient weight is obtained, a correction can be made
by constructing a curve as in Figure 11 of observed activity *vs.*
sample weight (specific activity constant) with sample thicknesses
less than "infinitely thick."

A word of caution should be inserted with regard to any counting
method that makes use of powdered or flaky material in such a way
that the active powder could flake off near the counter. This is
one of the easiest ways to raise the background of the counter and use
of such a method should be avoided where possible; if the method is
unavoidable, extreme care should be exercised. If at all possible the
material should be cemented or covered completely with paper. As
cement Zapon lacquer is good, and a dilute solution of Celluloid in
ethyl acetate–acetone or similar organic solvent is probably adequate.

4. Radioactivity Standards

For many research problems, absolute knowledge of the intensity
of the radioactive sources is unnecessary. However in some cases it
is desirable to know the total number of disintegrations coming from
the source. Although calibration methods using natural radioactive
materials, such as a weighed uranium oxide sample, have been stand-
ardized and can be used without insurmountable difficulties, a prefer-
able method is to obtain standardized sources from the National
Bureau of Standards (43). This bureau has available at a nominal
cost standards for β-ray counting as well as γ-ray standards, which
can be used for calibrating survey and monitoring instruments.
Radium and Co^{60} are both used for the latter purpose. In the meas-
urement of absolute activities, it is probable that the investigator will
need to determine the absorption characteristics of the element. Pro-
cedure for doing this will be found in texts on radioactivity and will
not be discussed further here.

F. UNIQUE PROBLEMS AND TECHNIQUES IN HANDLING RADIOACTIVE TRACERS

1. Radiochemical Problems

Though the subject of this chapter is the biological application of
radioactive tracers, there are certain topics in the chemistry of these
materials that should be mentioned. In handling radioactive materi-

als, particularly if little or no carrier is present, there are rules that must be taken into account and precautions that must be considered if the results are truly to mean what they are supposed to mean. When dealing with isotopes that are actually carrier-free one is working in a concentration range far below that of ordinary chemical experience.   For example, it can easily be calculated that a solution of $H_3P^{32}O_4$, completely free of stable P^{31}, with an activity of 1 mc. per liter will have a concentration of 10^{-10} M.   At such concentrations the tracer may exhibit a behavior quite different from that at ordinary concentrations.

In particular, one has to guard against removal of the tracer from solution by adsorption processes.   Even the small surface that a glass beaker presents to a solution contained in it may be enough to remove an appreciable fraction of the tracer from the solution.   The precipitation of an insoluble substance that is not supposed to carry the tracer is often likely to do so.   According to rules set down by men working in the field of natural radioactivity, such coprecipitation is only supposed to occur under certain definite conditions (44), but if the tracer is truly carrier-free it may and probably will be brought down when least expected.   For these reasons it is suggested that the use of completely carrier-free isotopes be avoided if it is possible to add enough carrier to bring the concentration of the solution to 10^{-4} M or greater.   If chemical operations at the lowest concentration range are necessary, each step should be examined carefully to make sure that the expected behavior is actually being obtained.   In general it is wise to be skeptical of precipitation processes or other steps in which a large surface is presented for adsorption.   Dry processes such as volatilization are safe from the standpoint of predictability, but every precaution must be taken to avoid the loss of the microgram or less of active material and the consequent contamination of the laboratory.   A completely closed system is mandatory for such operations.   The technique of liquid–liquid extraction is well adapted to the chemistry of carrier-free radioisotopes.   In most cases the behavior of high and low concentration levels is similar and no large surface is presented for adsorption.   Another technique that has had extensive study as applied to carrier-free substances is that of ion exchange.   In this case adsorption is utilized to aid in the desired chemical separation or process (45).

Though it may not seem likely, difficulty is frequently experienced in the simple process of diluting tracer with a carrier isotopic with it.

The assumption is always made that complete mixing occurs, but this is only true if the tracer and carrier are in the same chemical form or in forms that exchange rapidly. For example, iodine tracer in the form of free iodine will exchange rapidly with iodine in solution as iodide ion, but not iodate. It is frequently necessary to take the tracer and carrier together through several chemical processes involving valence changes before it can be assumed that they are actually mixed.

Once such mixing is obtained it may be desired to purify the tracer or at least to make certain that it is pure (45). In addition to the ordinary chemical operations that can be used for such purposes, including those listed above, certain special methods have been adapted for this purpose. One of these, which makes use of the adsorptive ability of some precipitates, is the procedure of "scavenging." Presumably carrier has been added only for the element desired and any radioactive impurity is still free of carrier or nearly so. A much larger fraction of the active impurity than of the desired element is then likely to be removed on an adsorbing precipitate. Care must be taken in choosing the precipitate. For example, in purifying strontium tracer from less basic elements such as yttrium or zirconium one could precipitate a hydroxide such as ferric hydroxide at a pH up to about 8 and remove large amounts of the yttrium or zirconium impurities by coprecipitation without affecting the stronitum. If the reverse purification (the removal of strontium from yttrium) were desired, however, one could not use this reaction since the yttrium would be more likely to coprecipitate than the strontium impurity. The beauty of radiochemistry is that the progress of such purification steps can be followed by comparing the characteristics of the activity separated on a scavenger precipitate with the main activity. Such comparison is important since the precipitate may just carry down some of the main activity and a single measurement of its radioactivity would not suffice to distinguish between impurity and product. Two general methods of comparison are by half-life and by absorptive characteristics. Since the important tracers usually have long half-lives, comparison on this basis would require long periods of time. Therefore, the absorptive characteristics, the ratio of β to γ activity, the shape of the absorption curve of the β rays in aluminum, or that of the X or γ radiation in aluminum or lead are frequently used for this identification. If such curves are identical for the scavenged and main activity it can be assumed either that the tracer is pure or that that particular reaction is not removing the

impurity. If they differ, repetition of the scavenging operation is indicated to insure removal of all the impurity (3, p. 115).

If the chemical nature of an active impurity is known, it is usually feasible to add that element in inactive form and make a chemical separation. The process of adding carrier for the active impurity and separating can be repeated as often as necessary. For example, if the chemical separation is only 80% effective, the first separation will leave 20% of the original impurity, but the second will leave only 20% of that or 4%, and so on. Again the radioactivity enables one to follow the process both qualitatively and quantitatively.

If the presence of an impurity of shorter half-life than the product activity is demonstrated, it may be possible to allow the impurity to decay out before using the tracer or before measuring the samples. Large differences in absorbability of the radiation may be utilized if the impurity has soft radiation and that from the product is hard. Measurement through an aluminum absorber of proper thickness will then cut out the soft radiation and permit the harder to pass through and to be a measure of the amount of the product activity present. If the energies of the radiations are reversed it may be possible to get the amount of product by difference probably with some loss in accuracy. A recent article on the subject of tracer purity is recommended as a source of further information (45).

In almost all biological work one deals with carrier and tracers that are isotopic. The chemical yield through any given steps is assumed to be identical to the yield of tracer or "radiochemical yield." Thus let us say that in a given purification process it is determined by weighing that an over-all chemical yield of 70% of the originally added carrier was obtained in the sample whose activity is measured. Then, dividing the activity of the sample by 0.7 is assumed to give the activity that would have been observed had the chemical recovery been 100%. In those few cases of biological interest in which isotopic carriers are not available for the tracer, e.g., 7.5 hour astatine (element 85), no correction can be made for chemical recovery. Thus, although iodine can be used as a carrier for astatine it cannot be assured that they behave identically, and indeed they do differ more than expected (46).

In summary, it must be said that before engaging in any large scale program in handling tracers through chemical operations, purifications, and other processes, a reference work in radiochemistry or, better, a qualified radiochemist should be consulted.

2. Special Problems in Using C[14]

The synthesis of biologically important compounds containing tracers is a research problem of considerable complexity. It is met most often in the case of organic compounds containing C^{14}, although the synthesis of organic sulfur, phosphorus, and iodine compounds with those elements labeled is also important. One of the principal difficulties in C^{14} synthesis (which applies equally well to syntheses involving C^{13}) is that of getting the active or labeling atoms in a proper compound as a starting point for the synthesis. For example, C^{14} was first made available as barium carbonate, which readily gives $C^{14}O_2$, but this is very difficult to use as a starting material. In addition, at a somewhat higher price, $C^{14}H_3OH$, $HC^{14}HO$ and some carboxyl-labeled fatty acids are now sold by the Isotope Division and other compounds will be available in the future. In addition, several commercial companies (*88,95,97*) have been authorized to prepare other intermediates and to sell them to institutions accredited by the Atomic Energy Commission. Correspondence with these companies or with the Isotope Division is suggested as the method of finding out what is available and its price.

Once the starting material is obtained, a method of synthesis that will label the compound in the proper position must be chosen with some care. For many problems it will be necessary to have the compound labeled in some specific position; for others, the tracer atoms can be scattered uniformly through the compound. Synthesis has been discussed in a number of recent review articles in which the experience using C^{13} has been applied to the essentially similar problems involved in use of C^{14} (*2*, p.161; *47*; *48*; see esp. *3b*). Some of the basic principles by which this differs from ordinary synthesis are: The reactant containing tracer is vastly more valuable than any other reactant and conditions are so adjusted to give maximum yield on the basis of this reactant. Dilution of the tracer is avoided to keep recovered tracer active enough for reuse. Adequate precautions are taken against accidental loss of the tracer, contamination of the equipment and laboratory, and overexposure of personnel. (Since one can be much surer of these last points if the system is completely closed, the use of a closed vacuum line type of apparatus is highly advisable.)

In the use of labeled organic compounds the possibility of exchange of tracer between its expected position and other molecules or other positions in the same molecule cannot be overlooked. Experiments

should be devised so that the chance of such exchange is minimized and interpretation of results should include a thorough examination of these possibilities.

Another example in which exchange can introduce error is in the samples prepared for measurement. If a solid such as barium carbonate is adopted as the form in which C^{14} will be measured, there is a chance that exchange will occur between the solid and atmospheric carbon dioxide. If such samples are stored for later measurement or checking, low results may be obtained. This problem has been investigated by several authors whose results appeared originally to be at variance (49,50). Study has shown that the original disagreement was caused by difference in mounting methods. It appears that a barium carbonate sample heated to 160°C. is fairly impervious to exchange, but a similar sample prepared just by filtering and drying may exchange up to 40% of its C^{14} in a few hours in an atmosphere of moist carbon dioxide. If it is inconvenient to prepare samples so that they can be heated to such a temperature they should be stored in the absence of atmospheric moisture or carbon dioxide. Other problems and suggestions are discussed by a number of authors (51,52).

3. Isotope Dilution Method

The possibility of tagging one kind of molecule by radioactive or stable tracer has given rise to a new method of analysis, the isotope dilution method (53). In analyzing a complex mixture of organic compounds—take, for example, the amino acids liberated on hydrolysis of a protein—it is relatively easy to separate out a particular amino acid in pure form, but much more difficult to obtain the quantitative yield that would be necessary for analysis. If a particular amino acid is tagged, say with C^{14} in some nonexchanging position, added in known amount to the mixture in such a way that complete mixing is obtained, and then separated by the procedure that is standard for that amino acid, the separated amino acid will have a lower specific activity of C^{14} than that added. The amount of reduction in the specific activity, or the dilution, is directly related to the amount of the amino acid originally present in the mixture.

Samples of the added and separated amino acids could be measured directly or one could burn them to carbon dioxide, absorb this in barium hydroxide, and measure the properly mounted barium car-

bonate precipitates. In either case the following relations between weights and specific activities would be true: If X_1 and X_2 are the specific activities of the added and separated samples and W_1 and W_2 are the weights of amino acid added and originally present in the mixture, the relation:

$$\frac{X_1}{X_2} = \frac{W_1 + W_2}{W_1}$$

will hold. Solving this for W_2, the unknown weight, we obtain:

$$W_2 = W_1 \frac{(X_1 - X_2)}{X_2}$$

If the weight of added material W_1 can be neglected in relation to W_2 or for a tracer available in some carrier-free form that will mix with the substance in the mixture, the product W_1X_1 can be replaced by the observed total activity, A_1, of tracer added: the expression then simplifies to $W_2 = A_1/X_2$.

4. Radioautographs

The radiations given off by radioactive isotopes permit a unique scientific use of these isotopes, in that their position in tissue can be located merely by allowing them to take their own picture. Such a picture is called a radioautograph or autoradiograph. The method was first applied to the gross location of the element in the plant or animal being studied. Some examples of these are the studies by Stout and others (54) of the uptake of phosphorus and of certain trace elements in tomato plants (Fig. 12), which show, among other things that concentration occurs in the seeds of the tomato. Other examples of the technique are the studies using this method that have given better understanding of the plant circulatory system.

The real possibility of the method lies in its application to microscopic studies. A radioautograph prepared in the usual way may on microscopic study yield information about the location of the tagged atoms within the individual cells. While most tracers can be used for the preparation of macroradioautographs, application in microscopic study places some restrictions on the usable isotopes. The darkening of the photographic plate is, of course, restricted to the paths of the various particles. Since the particles differ widely in their penetrative power the length of their paths in the photo-

graphic emulsion will also differ, and the longer this path, the more diffuse will be the spot on the photograph corresponding to a point of location of a radioactive element in the original.   For this reason α particles with their short ranges and high energies give the best radioautographs (3, p. 130).   If α emitters were biologically important they would find wide use for this purpose; since they aren't, it is necessary to use the β-ray emitters that are important and to be satisfied with more diffuse autographs.   An element giving only X or
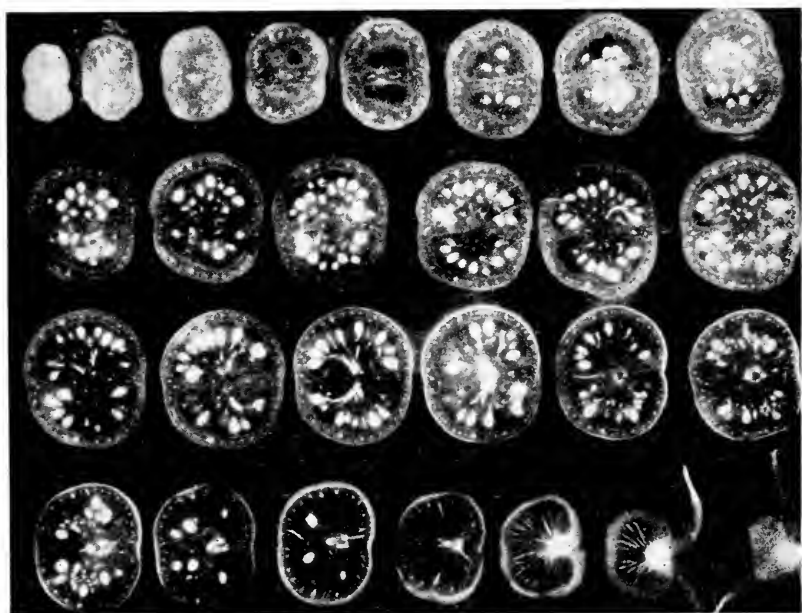
Fig. 12.   Radioautograph of zinc in a single tomato.   This technique makes visible 3×10^{-9} g. zinc present in each tomato seed.   (Courtesy P. R. Stout.)

γ radiation, or a very high proportion of such would be useless from a radioautographic point of view.

Also with a view to obtaining the least possible diffuseness and the highest possible resolution, one must use tissue slices that are as thin as possible.   Best success has been obtained with a thickness of 5μ, though 10μ has often been used.   Contact between the tissue section and the photographic emulsion should be as intimate as it can be made.   If direct contact cannot be made, intervening material should be thin; a film of celloidin or similar material has been used.

One of the principal difficulties is the correlation of the radio-autograph with the microscopic structure of the material. It is obviously necessary that correlation be made if the method is to be of value. The use of fiducial marks on the tissue section and the photograph is one way of obtaining such correlation, but with anything but low magnification, this method is not capable of giving good register. It should be borne in mind that the tissue itself usually must be stained before histological study can be made. Several techniques have been experimented with in which the tissue slice and photographic emulsion are not separated after exposure of the latter. In one method the tissue slice that contains tracer is mounted on the photographic plate or film which is then set aside for the proper exposure period, the radioautograph is developed and fixed and the tissue is treated with the proper histological stain (56). In another technique the section is mounted on a microscope slide and emulsion in liquid form is poured over it. Exposure and processing of the photograph are done in the usual manner. Staining in this case is more difficult since the section is covered by the emulsion, but staining before the exposure is not a sure procedure since the radioactive atoms would be likely to move during the staining procedure.

To go into the subject beyond this brief statement is beyond the scope of this chapter but the following references to excellent articles will help the beginner (3, p. 122; 54a,54b,57–59).

G. MONITORING RADIATION FOR HEALTH AND SAFETY

Closely related to the problem of measurement is that of monitoring or personnel protection. In order that the neophyte in radiotracer research can have an appreciation of the subject and an understanding of the terms used in its discussion, a brief portion of this chapter is devoted to the subject.

1. Calculations

The accepted unit of radiation dose in tissue is the roentgen or r. unit (see chapters by Gowen and by Gray) and others are derived from it (17). It corresponds to an energy dissipation in air of 84 ergs per gram. Hence the other commonly used units for dosage measurements, the roentgen equivalent physical (r.e.p.) or equivalent roentgen (e.r.), are defined as the quantity of ionizing radiation of any type that may be absorbed at the rate of 84 ergs per gram. Though

derived for air, the same energy absorption per gram is used for tissue as well.   For many purposes the two units can be considered synonymous.   One thousandth part of each of these units, the milliroentgen (mr.) and millirep (mr.e.p.) are frequently used.   Another unit recently introduced, the r.h.m. (roentgens per hour at one meter), is more for the comparison of source strengths than dosages (17).

From the experience of roentgenologists and radium workers it has been decided that the amount of radiation that can be absorbed by the body each day without damage is 100 mr.   This so-called "tolerance value" represents an absolute maximum if the worker is subjected to it each day and a working maximum if the exposure is infrequent.* Since genetic effects probably occur at lower levels than those that produce observable damage, it is well to consider 100 mr. per day as a maximum even for occasional exposure and to keep the average exposure to less than 10 mr. per day.   On the basis of an eight hour working day, tolerance is frequently given as 12.5 mr. per hour. Although these levels were originally proposed for radiation received by the whole body it is safer to consider 100 mr. as a maximum daily dose for any part of the body (2, pp. 395, 409).

The estimation of expected radiation dosage from external sources under given conditions of source strength, type and energy of radiation, distance and shielding is not a particularly difficult problem. It will differ for each kind of radiation but γ rays constitute the only real hazard from external sources.   Since several of the quantities involved are known only approximately the calculation cannot be made with any great accuracy, but such accuracy is not necessary for the purpose, results within 20% being adequate.   To this accuracy the absorption coefficient of γ rays in air (μ_a) can be assumed constant over the range 0.05 to 2 m.e.v. and the value 3.6×10^{-5} cm.$^{-1}$ is used. Below this range the gamma rays are easily shielded and no important tracers have γ rays with energy appreciably above 2 m.e.v.

The formula for calculating the dosage rate in roentgens per hour from a source of m millicuries at distance d cm. can then be put together as follows:

$$\text{Dosage rate} = \frac{E e \mu_a A}{W 4\pi d^2} = \frac{5.6 E m}{d^2} \text{ r./hr.}$$

$E = \gamma$-ray energy in million electron volts, $e =$ charge on ion pair $= 4.8 \times 10^{-10}$ e.s.u., $A =$ disintegrations per hour $= 3.7 \times 10^7 \times$

* Recently, many laboratories have reduced their operating limit to 50 mr./day.

$3600 \times m$, and W = energy required to produce one ion pair = 32.5×10^{-6} m.e.v. in air.

The uncertainties in W and μ make this equation only an approximation. As written it applies only if each decay process gives one and only one γ ray. For a substance with a complex decay scheme it is necessary to replace E in the above by $\sum_i p_i E_i$ in which p_i is the fraction of the total number of disintegrations in which a γ ray of energy E_i is given off.

In order to calculate the amount of shielding necessary to bring the radiation at a given distance down to the desired value, the half-thickness of the radiation must be obtained from a curve such as Figure 2. From this half-thickness the necessary shielding can be calculated or one can determine the safe working distance with given shielding (60). Calculation of dosage for β radiation is less certain. β rays do not have enough penetration to constitute a serious hazard if shielding is provided in the walls of the containing vessel or in the gloves and clothes of the worker. Only in the case of high energy β-ray emitters will additional protection be necessary and it is more readily provided by heavy-walled apparatus than by a shield.

The same principles apply if a worker, experimental animal, or biological system is exposed to radiation from an internal source, whether by accident or design. Again one assumes maximum tolerance to be 100 mr. per day for γ radiation, but now the day is 24 hours long. Tolerances for other types of radiation have been set at 100 mr.e.p. for β rays and 10 mr.e.p. for α particles. Calculations in this field involve a knowledge of the behavior of the particular element in the body and its possible concentration in the various organs, its biological elimination rate and other factors. For these aspects of the problem, the reader is referred to the chapter by Gray as well as to several recent articles (62,63).

2. Instruments and Methods

Monitoring assumes several aspects. Personnel must be monitored to see that they do not get overdoses while handling the active material. Surfaces of working areas, apparatus, and glassware must be surveyed after active materials have been handled and possibly spilled to reduce the possibility of ingesting and inhaling radioactive dust and, more serious, to eliminate the danger of contaminating one experiment with the spilled activity from previous ones. Personnel

monitoring requires measuring devices that can be worn or carried
by the worker without interfering with his manipulations.   One of the
surest means of determining the radiation dosage that a worker has
received is the photographic badge.   The film badge technique in-
volves a considerable knowledge of procedure in handling the film and
in preparing standards that the ordinary laboratory is not likely to
have or to wish to acquire.   It is gratifying that the isotopes distri-
bution program has been expanded to include on request a film badge
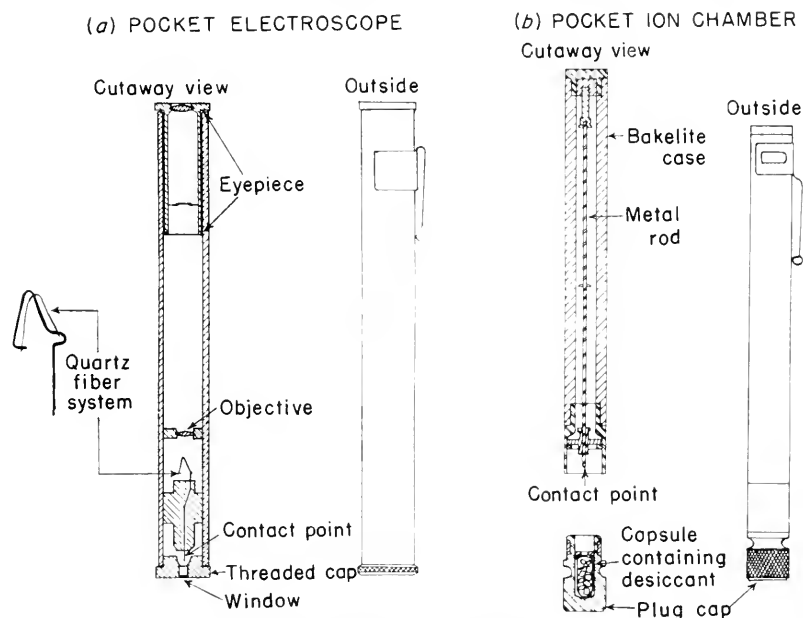
Fig. 13.   Several types of pocket radiation dosimeters.

service for laboratories to which isotopes have been sent (*19*).   This
cannot be said to give adequate protection since one knows about the
exposure only after it has been received.   However, it does constitute
a worthwhile safeguard, since a person who has received an excess
dose as indicated by the badge could be relieved of such duty, or the
work could be revised to reduce the hazard.   Some development has
been done on ring film holders to be worn on the hand likely to receive
a much greater dose than the remainder of the body; however, this
phase of the art is not in as advanced a state as that of film badges.
   Several types of ionization chambers that can be worn by the

worker are available.   The older of these is a chamber made of Bake-
lite or other plastic about the size of a fountain pen (Fig. 13b) (*84,87,
92*).   The inside of the chamber wall is coated with an electrically
conducting film.   This film and a thin metal rod down the center, but
insulated from this coating, constitute the two electrodes.   The
chamber is charged by plugging it into an a.c.-operated electrometer,
which makes contact with the two electrodes.   It is read also by
means of this electrometer (*64*).

Another type of pen-sized pocket chamber was one developed
during the war by O. G. Landsverk and now made by a number of
instrument makers.   This is truly a pocket electrometer since each
chamber contains its own quartz fiber and optical system and oper-
ates on the general principle of the Lauritsen electroscope described
earlier (Fig. 13a).   It is charged by a battery-operated charging box,
which also allows one to view the fiber during the charging operation.
Its principal advantage is the fact that it can be viewed at any time
by merely looking through it at a light source (*83,84,92*).   The sim-
pler chamber is somewhat less accurate, less convenient, a little more
rugged, and considerably less expensive than the latter.   Either
should do an adequate job of on-the-spot personnel monitoring for
most laboratories, though at least a few pocket electrometers are
recommended for medium to high level work since the worker can
tell very quickly whether his radiation dose rate is approaching the
dangerous level.

Of equal importance are the nonwearable but still portable survey
meters.   These are battery operated and usually weigh 5–10 pounds.
As usual two types are available: GM tube and ionization chamber
survey meters.   In the first type the integrated current of the tube
pulses or, in the second, the chamber current is amplified and made to
operate a direct current meter.   Since the dose is dependent on the
amount of ionization and not solely on the number of particles, it is
expected that the ion chamber meter, which measures the actual
ionization, will be more accurate than the GM tube, which merely
counts the particles.   However, experience favors the GM tube sur-
vey meters for their stability, ease of handling and repair, and more
rapid response to a change in radiation intensity.   The circuits of
the ionization chamber survey meters are frequently slow in respond-
ing, having time constants of five to twenty seconds.   The GM tube
survey meter responds more rapidly and is usually equipped with ear
phones through which the actual pulses are fed, producing an in-

stantaneous clicking noise.   This lack of delay is of advantage in
that survey measurements can be made much more rapidly if the
meter responds rapidly.   In general, choice between types of instru-
ments depends on the particular application.   The GM tube meters
should, for accurate work, be calibrated with sources of different kinds
and energies of radiation, while the ionization chamber type gives a
more accurate measurement of the ionization producible by the radia-
tion in air or in tissue.

The GM tube type can be made more sensitive than the ion cham-
ber type.   Currently available GM tube survey meters are provided
with three scales: 0.2, 2, and 20 mr. per hour full scale (82,84,87).
Since, in an eight hour day, tolerance is 12.5 mr. per hour these meters
cover the range between tolerance and background, which is about
0.01 to 0.02 mr. per hour in the absence of all radioactive contamina-
tion.

The ion chamber type of survey meter is another wartime develop-
ment originally designed for monitoring contamination by α-particle
emitters.   Later modifications were developed capable of measuring
any radiation and were provided with screens to permit discriminating
measurement of the different types of radiation.   Several of these are
now available commercially (87,93,94), having sensitivity scales
ranging from 2 to 2000 mr. per hour full scale.   Other types will prob-
ably be available in the near future.   Of those made available re-
cently several models with fairly good sensitivity toward all radiation
(93) and for γ rays only (87) have proved satisfactory.

There are several a.c.-operated, nonportable instruments that
give a direct meter reading of the radiation incident upon the ion
chamber.   These can be made somewhat more accurate than the
portable models but they are useful only for a qualitative or semi-
quantitative measure of the strength of sources since an accuracy of
3–5% is all that can be expected from a reading on the meter dial
(82,85).   Experience shows that the instruments should be cali-
brated;   this can be done by using the formula developed in Section
G1 if strong enough γ sources ($\sim$ 1 mc.) are available.   Operating
manuals that are helpful for a technician moderately experienced in
electronics are furnished by the manufacturers.

II. PLANNING A RADIOCHEMICAL LABORATORY

The potential hazards involved in handling radioactive materials
necessitate serious thought about the construction of the laboratory in

which they are to be used.   Since the widespread use of tracers is just beginning there is no large backlog of experience to guide the designers of laboratories adapted for this purpose in the various fields of biology, medicine, agriculture, and so on, in which tracers will be applied.   Interest in the subject resulted in the symposium on the design of radiochemical laboratories at the April, 1948, American Chemical Society meeting (65).   The difficulty with this and other such
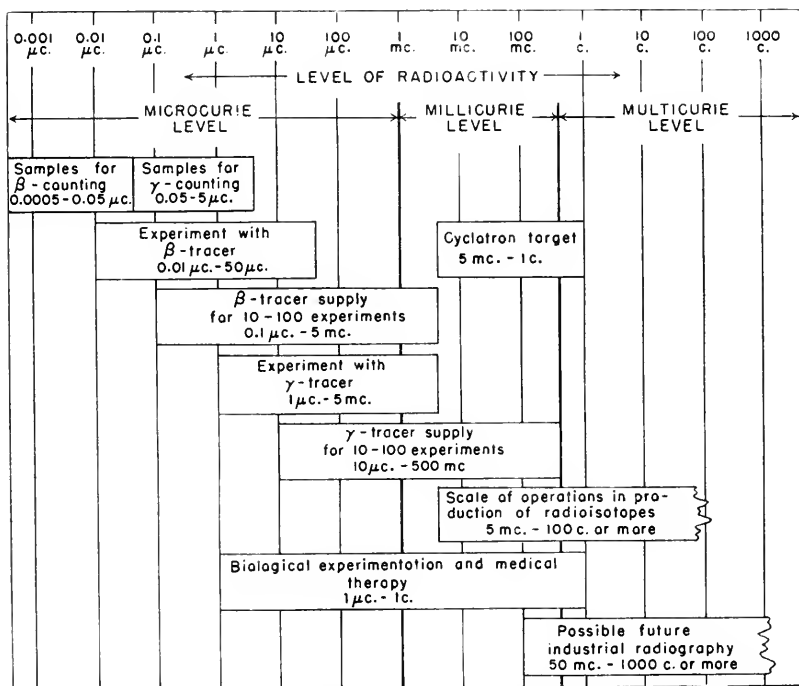
Fig. 14.   Approximate range of radioactivity encountered in various problems (66)

presentations is that they have been largely discussions of practice in Atomic Energy Commission laboratories, in which the activity levels have been far above those of the ordinary tracer laboratory and for which funds are available to a degree unheard of for the small research or educational institution.   In fact, it should be realized at the outset that building and equipping a tracer laboratory may be considerably more expensive than is the case for laboratories for other purposes (66,67).

Opinions on the design of laboratories and the extent of precautions

necessary are likely to differ and to change as more experience is obtained. At the present time a national body, the National Committee on Radiation Protection (68), and sectional bodies such as one representing the thirty Midwest institutions that co-operate with the Argonne National Laboratory are drafting regulations on the safe handling of radioisotopes, including laboratory design (61). Since these regulations are not in a final state anything said now may become obsolete. Eventually some such body of regulations will apply to anyone desiring to work in this field. The consideration that needs to be given to laboratory design is highly dependent upon the nature and
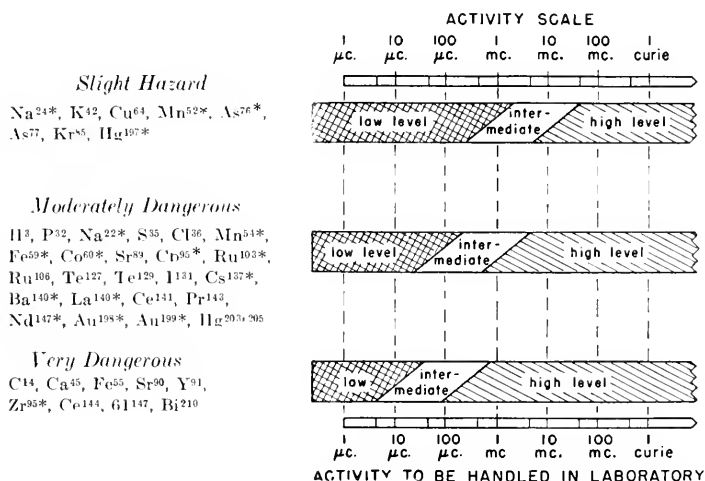
Fig. 15. Toxicity levels of selected radioisotopes (68).

Effective radiotoxicity is obtained from a weighting of the following factors: (a) half-life, (b) energy and character of radiations, (c) degree of selective localization in the body, (d) rates of elimination, and (e) quantities involved and modes of handling in typical experiments. The slant boundaries between levels indicate borderline cases and emphasize that there is no sharp transition between the levels and the associated protection techniques. The principal γ-ray emitters are indicated by asterisks (e.g., Na^{24}*).

amount of the particular isotopes to be handled, in addition to the type of work to be done. The amounts of activities of various sorts that will be required for different operations and processes are shown in Figure 14 (66).

The provisional bulletin of the National Committee on Radiation Protection (68) lists isotopes in three categories of danger, depending largely upon half-life, but also on radiation type and energy and biological behavior (Fig. 15).   Thus if a laboratory is to be restricted to the low level region of this table, no particularly different features need be built in or provided other than good ventilation.   At the intermediate level more thought will have to be taken in the design and for high level work special design is almost imperative.   Since this table is presented on the basis of internal radiation it does not cover the danger from the external radiation of γ-ray emitters.   The dosage from the latter and the necessary shielding can be calculated as shown earlier.

Many operations in the usual tracer laboratory can be handled with temporary shielding, preferably lead because of its density.
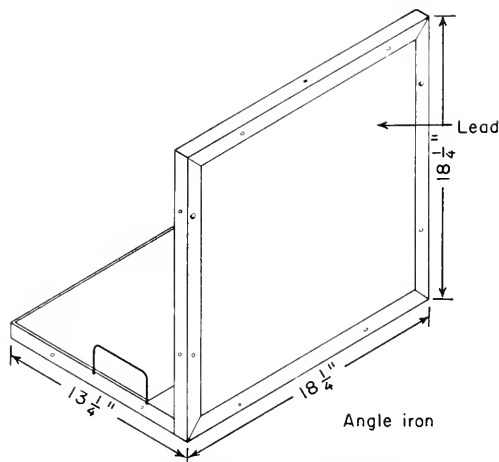
Fig. 16.   A useful semiportable lead screen.

Frequently size $2 \times 4 \times 8$ in. lead bricks are used since these possess a convenient weight (about 25 lb.) and shape.   The main difficulty with such bricks is that a 2 in. wall built with them is not very stable and a 4 in. width is in many cases unnecessary.   A semiportable lead screen of L shape made of 1 in. lead bolted to an angle iron frame has proved to be very useful for handling those isotopes whose γ rays have energies less than 0.7 m.e.v. (Fig. 16).

The disposal of radioactive wastes is a serious problem.   A large number and several types of substances must be considered.   One of

the most serious is $C^{14}O_2$ which may occur as the respiration product of animals treated with C^{14} in some organic compound, or in some chemical operation. Though this may prove difficult it is considered much better practice to try to capture and retain such C^{14}, if present in amounts greater than a few microcuries, than to dilute it and disperse it. It is because of such problems as these that the apparent advantage of C^{14} over C^{13} as a tracer on closer observation becomes less clear-cut. Gaseous wastes from other tracers may include I^{131} as iodine or hydrogen iodide, S^{35} in various forms, and any number of organic molecules containing other tracers. The maximum level of activity set by the National Committee (68) is 10^{-5} μc. per liter of air.

Chemical operations such as organic syntheses involving C^{14} (and other tracers) can be performed in a closed system with much more safety than in an open system. Thus any $C^{14}O_2$ or other radioactive gas can be absorbed in some chemical trap rather than disposed to the surroundings. With liquid wastes the problem again depends on the nature of the isotope. For short-lived materials (half-life less than one month), storage of active solutions as such until decayed is highly feasible and is probably the best answer. With those of longer half-life it is best to precipitate the activity in the form of some insoluble compound, and store that, being sure to check the activity of the supernatant before discarding it into the drain. In no case should active solutions be allowed to get into a public drain without monitoring or checking of some sort. The level of activity that is safe to pour into a drain is another matter of controversy and any existing regulations or suggestions extant at the time should be checked. Some laboratories may be fortunate enough to be situated so that an auxiliary drain can be built to carry the active liquid waste to a point from which it can under no circumstances get into a public water supply. Such a drain may safely carry a somewhat higher level of activity but the possibility of concentration of the radioactivity by plant life to a dangerous point puts a fairly low limit on the maximum quantity disposable in this way. If a new laboratory is being built or if one has a possibility for remodelling, a tank to contain the possibly radioactive liquid wastes to allow for monitoring should be installed.

On the subject of radiochemical laboratory design, the only real conclusion that can be reached at present is that the subject is not static and that the suggestions and regulations that appear in the literature between the date of this writing and the time this is being

read may supersede or make obsolete any specific suggestions made
here.   One thing is clear, it is much better to get experience with
tracers by handling the safe ones at low levels first and after that to
progress to higher levels and more dangerous isotopes (*3*, p. 169;
69).

I. BIOLOGICAL APPLICATIONS OF RADIOACTIVE TRACERS

To discuss, even briefly, the vast number of applications of radio-
active tracers to biological problems that have appeared in print to
this date would require a chapter much longer than the present one.
It is not possible in all cases to draw a line between stable and radio-
active tracer research.   Thus, in the case of carbon, early tracer
work with C^{11} and continuing work with C^{13} are intimately associated
with the increasing amount of work using C^{14}, now that it is widely
available.   Research using tracers is the subject of a number of re-
views published recently to which readers are referred for additional
information and references to the original literature (*70,71*).

In the case of the hydrogen isotopes, very little has been done with
tritium because it is unavailable and difficult to measure.   The work
with deuterium is reviewed in Chapter XVI on stable isotopes.

There is obviously more interest in tagging carbon atoms than
any others and the availability of the several carbon tracers has al-
ready led to increased knowledge in a number of biological problems.
The basic problems of photosynthesis and metabolism have duly re-
ceived a large share of this research (*1*, p. 148; *2*, p. 182 ff., *3b,72–75*).
References to some of the survey articles are given in the preceding
chapter since much of this research was done with C^{13}.   Some idea
of the amount of work accomplished and that yet to be done can be
obtained from these reviews.   Typical of other uses of radiocarbon
are the studies of individual biological reactions, as for example that
showing that phenylalanine is the precursor of epinephrine (*76*).

One of the most important radioactive tracers is P^{32} by reason of
its half-life and other characteristics as well as its biological impor-
tance (*1*, p. 184; *2*, pp. 209, 261; *71*).   Since many metabolic processes
involve phosphorus in the form of phosphate esters, P^{32} has been
widely used in the study of metabolism of carbohydrates, phospho-
lipides, and nucleic acids.   Other studies have dealt with other en-
zymic reactions, with the mechanism of insulin action, and with the

specific behavior of phosphates in the body in normal and diseased animals. Due to the importance of phosphates in plant growth, plant physiological and agronomic studies using P^{32} have assumed large scale proportions.

The alkali metal tracers Na^{24} and K^{42} and the halogen tracers Cl^{38} and Br^{82} have been used to study the permeabilities of cell boundaries and the rate of transfer of salts across them (1, p. 213; 2, p. 261; 77,78). By measuring the extent of dilution of the tracer the volume of extracellular fluid has been determined. The existence of sodium in bone, which does not exchange rapidly with plasma sodium, has been shown with Na^{24}. Experiments with ground bone showed this retention to be an adsorption process.

Studies with other isotopes have largely been concerned with the biological behavior of the individual elements and the important compounds of which they are an integral part. For example, S^{35} has been used in studies on the sulfur-containing amino acids, cystine and methionine (1, p. 201; 79), their interrelations, metabolism, and so on. Fe^{59} has been used in studies of iron usage and processes involving hemoglobin (1, p. 235; 80); by biological synthesis it has been incorporated into erythrocytes, which have then been used in studies on the fate of red cells, on blood volume, etc. Ca^{45} and a rather inadequate substitute, Sr^{89}, have been used to study such topics as calcium metabolism, processes in bone formation, and the like.

Of particular interest are studies using I^{131}, most of which have, obviously, been related to the functioning of the thyroid gland. The mechanism of thyroid action in concentrating iodine and converting it into diiodotyrosine and thyroxine, the malfunctioning of the thyroid in natural and drug-induced pathogenesis, and similar topics have been widely studied (2, p. 292).

A completely distinct field of tracer application is that of medical use (2, p. 327). Anyone requesting isotopes for human use is required to satisfy a special committee set up by the National Research Council and the Atomic Energy Commission that he is capable and understands the problems involved. This requirement makes it less likely that anyone with insufficient training will obtain isotopes for such a purpose. One application that can be included in this category is that of diagnosis, of which typical examples follow. In amputations because of gangrene, Na^{24} has been injected into the blood stream in order to determine with an external counter the extent of the active circulation in the area of the proposed amputation to fix the position

of the latter as accurately as possible. I^{131} has had a number of diagnostic applications. For example, since the rate of iodine uptake is greater in the case of hyperthyroidism, this condition can be diagnosed by the use of I^{131} and an external counter. Another application of I^{131} not related to the thyroid has been reported by Moore (*81*), based on the fact that fluorescein and its derivatives are preferentially taken up by tumor tissue. Diiodofluorescein containing I^{131} was injected into patients previously diagnosed as having brain tumors. In the relatively few cases reported the physician was able to determine by an external counter the existence and location of the suspected tumor. In all cases in which an operation was performed the counter diagnosis was found correct.

The other application of radioisotopes in medicine is, of course, therapy. Radioisotope therapy requires activities several orders of magnitude greater than those used in tracer work and should be carried out only by institutions that include on their staffs radiologists familiar with the literature and capable of calculating proper dosages. The substitution of Co^{60} or other long-lived isotope in a capsule or wire for the usual radium or radon needle in the irradiation of cancer tissue is not really a different technique but a modification of the old. The really new technique is the use of a radioisotope that will by reason of its chemical nature (or the nature of the compound of which it is made a part) locate itself at the site of the malignancy. Some of the principal successes of artificial tracers in this field are the use of P^{32} for the treatment of polycythemia vera (*2*, p. 353) and of I^{131} for inoperable hyperthyroidism (*2*, p. 377). A large fraction of the total shipment of isotopes from Oak Ridge has gone to hospitals that make such use of them. It is a field somewhat foreign to the subject of this chapter since it is governed by the rules of medicine and not those of health physics. In other words, in radium, radon, X-ray, or artificial isotope treatment of cancer or other malignant conditions, doses are given to the patient that are as much as a thousand to ten thousand times tolerance. Such doses are allowable only as the lesser of two evils, the other being the death of the patient if the pathological condition is not checked.

References

GENERAL

1. Kamen, M. D., *Radioactive Tracers in Biology*. Academic Press, New York, 1947.

2. *Symposium on the Use of Isotopes in Biology and Medicine.* University of Wisconsin Press, Madison, 1948.

3. U. S. Naval Medical Bulletin, 1948, supplement to March–April issue.

3a. Hevesy, G., *Radioactive Indicators.* Interscience, New York, 1948.

3b. Calvin, M., C. Heidelberger, J. C. Reid, B. M. Tolbert, and P. E. Yankwich, *Isotopic Carbon.* Wiley, New York, 1949.

3c. Lawrence, J. H., and J. G. Hamilton, *Advances in Biological and Medical Physics.* Academic Press, New York, 1948, Vol. I.

3d. Symposium: *Nucleonics and Analytical Chemistry.* Evanston, Ill., Aug., 1948; *Anal. Chem.,* **21,** 318 (1949).

GENERAL RADIOACTIVITY

4. Cork, J. M., *Radioactivity and Nuclear Physics.* Van Nostrand, New York, 1946.

5. Rasetti, F., *Elements of Nuclear Physics.* Prentice-Hall, New York, 1936.

6. Hoag, J. B., and S. A. Korff, *Electron and Nuclear Physics.* Van Nostrand, New York, 1947.

7. Pollard, E., and W. L. Davidson, *Applied Nuclear Physics.* Wiley, New York, 1942.

7a. Lapp, R. E., and H. L. Andrews, *Nuclear Radiation Physics.* Prentice-Hall, New York, 1948.

7b. Friedlander, G., and J. W. Kennedy, *Introduction to Radiochemistry.* Wiley, New York, 1949.

8. Seaborg, G. T., *Chem. Revs.,* **27,** 199 (1940).

9. Hevesy, G., and F. A. Paneth, *A Manual of Radioactivity.* Oxford Univ. Press, London, 1938.

10. McMillan, E. M., E. Segrè, *et al., Lecture Series on Nuclear Physics.* (MDDC-1175), U. S. Govt. Printing Office, Washington, 1947.

INTRODUCTION

11. Hevesy, G., *Biochem. J.,* **17,** 439 (1923).

12. Hevesy, G., *Ann. Rev. Biochem.,* **9,** 641 (1940).

13. Manhattan Project, *Science,* **103,** 697 (1946).

14. Isotopes Division, USAEC, *Science,* **106,** 175 (1947).

BASIC RADIOACTIVITY

15. Kohman, T. P., *Am. J. Phys.,* **15,** 356 (1947).

16. Glendenin, L. E., *Nucleonics,* **2,** No. 1, 12 (1948).

17. Evans, R. D., *Nucleonics,* **1,** No. 2, 32 (1947).

18. Condon, E. V., and L. F. Curtiss, *Science,* **103,** 712 (1946).

AVAILABILITY OF TRACERS

19. Isotopes Division, USAEC, *Radioisotopes, Catalog and Price List, No. 2.* Oak Ridge, Tenn., Sept., 1947.

20. Szilard, L., and T. A. Chalmers, *Nature*, **134,** 462 (1934).

21. Seaborg, G. T., and I. Perlman, *Rev. Modern Phys.*, **20,** 585 (1948).

22. *Trilinear Chart of Isotopes*, W. H. Sullivan, ed., Wiley, New York, 1949.

INSTRUMENTATION

23. Korff, S., *Electron and Nuclear Counters*. Van Nostrand, New York, 1946.

24. Lewis, W. B., *Electrical Counting*. Macmillan, New York, 1943.

25. Lauritsen, C. C., *Rev. Sci. Instruments*, **8,** 438 (1937).

26. Corson, D. R., and R. R. Wilson, *Rev. Sci. Instruments*, **19,** 207 (1948).

27. Simpson, J. A., *Rev. Sci. Instruments*, **18,** 884 (1947).

28. Neher, H. V., in Strong, *Procedures in Experimental Physics*. Prentice-Hall, New York, 1943, pp. 259 *ff*.

29. Neher, H. V., and W. W. Harper, *Phys. Rev.*, **49,** 940 (1936).

30. Copp, D. H., and D. M. Greenberg, *Rev. Sci. Instruments*, **14,** 205 (1943).

31. Brown, S. C., *Nucleonics*, **2,** No. 6, 10; **3,** No. 2, 50; No. 3, 46 (1948).

32. Higinbotham, W. A., J. Gallagher, and M. Sands, *Rev. Sci. Instruments*, **18,** 706 (1947).

33. Libby, W. F., *Ind. Eng. Chem., Anal. Ed.*, **19,** 2 (1947); *Phys. Rev.*, **46,** 196 (1934); **55,** 245 (1939).

34. Henriques, F. C., Jr., and C. Margnetti, *Ind. Eng. Chem., Anal. Ed.*, **18,** 417, 421 (1946).

35. Miller, W. W., *Science*, **105,** 129 (1947).

36. White, L., Jr., *Determination of Radioactive Carbon*. Isotopes Division Circular A2, USAEC, Oak Ridge, Tenn., 1947.

37. Henriques, F. C., Jr., G. B. Kistiakowsky, C. Margnetti, and W. G. Schneider, *Ind. Eng. Chem., Anal. Ed.*, **18,** 349 (1946).

38. Bale, W. D., and J. F. Bonner, Jr., in A. Weissberger, ed., *Physical Methods of Organic Chemistry*. 2nd ed., Interscience, New York, 1949.

39. Solomon, A. K., and H. D. Estes, *Rev. Sci. Instruments*, **19,** 47 (1948).

ERRORS AND CORRECTIONS

40. Reid, A. F., A. S. Weil, and J. R. Dunning, *Ind. Eng. Chem., Anal. Ed.*, **19,** 821 (1947).

41. Rainwater, L. J., and C. S. Wu, *Nucleonics*, **1,** No. 2, 60 (1947); **2,** No. 1, 42 (1948).

42. Yankwich, P. E., T. H. Norris, and J. Huston, *Ind. Eng. Chem., Anal. Ed.*, **19,** 439 (1947).

43. National Bureau of Standards, *Nucleonics*, **1,** No. 4, 63 (1947).

SPECIAL PROBLEMS

44. Hahn, O., *Applied Radiochemistry*. Cornell Univ. Press, Ithaca, New York, 1936.

44a. Schweitzer, G. K., and I. B. Whitney, *Radioactive Tracer Techniques*. Van Nostrand, New York, 1949.

45. Cohn, W. E., *Anal. Chem.*, **20**, 499 (1948).

46. Corson, D. R., K. R. MacKenzie, and E. Segrè, *Phys. Rev.*, **58**, 672 (1940).

47. Price, T. D., and W. W. Miller, *Nucleonics*, **1**, No. 3, 4; No. 4, 11 (1947).

48. Loftfield, R. B., *Nucleonics*, **1**, No. 3, 54 (1947).

49. Yankwich, P. E., *Science*, **107**, 681 (1948).

50. Armstrong, W. D., and J. Schubert, *Science*, **106**, 403 (1947).

51. Dauben, W. C., J. C. Reid, and P. E. Yankwich, *Ind. Eng. Chem., Anal. Ed.*, **19**, 824 (1947).

52. Armstrong, W. D., and J. Schubert, *Anal. Chem.*, **20**, 270 (1948).

53. Henriques, F. C., Jr., and C. Margnetti, *Ind. Eng. Chem., Anal. Ed.*, **18**, 476 (1946).

RADIOAUTOGRAPHS

54. Arnon, D. I., P. R. Stout, and F. Sipos, *Am. J. Botany*, **27**, 791 (1940).

54a. Glick, D., *Techniques of Histo- and Cytochemistry*. Interscience, New York, 1949, pp. 152–158.

54b. Yagoda, H., *Radioactive Measurements with Nuclear Emulsions*. Wiley, New York, 1949.

55. Pecher, C., *Univ. Calif. (Berkeley) Pubs. Pharmacol.*, **2**, 117 (1942).

56. Evans, T. C., *Radiology*, **49**, 206 (1947).

57. Evans, T. C., *Nucleonics*, **2**, No. 3, 52 (1948).

58. Gorbman, A., *Nucleonics*, **2**, No. 6, 30 (1948).

59. Belanger, L. F., and C. P. Leblond, *Endocrinology*, **39**, 8 (1946).

HEALTH AND SAFETY

60. Morgan, G. W., *Gamma and Beta Radiation Shielding*. Isotopes Division Circular B-3, USAEC, Oak Ridge, Tenn., 1948.

61. Clinton Laboratories, USAEC, *Nucleonics*, **1**, No. 4, 60 (1947) (general rules and procedures concerning radioactive hazards).

62. Morgan, K. Z., *J. Phys. & Colloid Chem.*, **51**, 984 (1947).

63. Marinelli, L. D., E. H. Quimby, and G. J. Hine, *Nucleonics*, **2**, No. 4, 56; No. 5, 44 (1948); *Am. J. Roentgenol. Radium Therapy*, **59**, 260 (1948).

64. Rinaker, R. E., *Nucleonics*, **2**, No. 1, 78 (1948).

65. *Ind. Eng. Chem.*, **41**, 227–50 (1949).

66. Levy, H. A., *Chem. Eng. News*, **24**, 3168 (1946).

67. Sullivan, W. H., *Chem. Eng. News*, **25**, 1862 (1947).

68. *Safe Handling of Radioisotopes*. National Committee on Radiation Protection, 1948.

69. Henriques, F. C., Jr., and A. P. Schreiber, *Nucleonics*, **2**, No. 3, 1 (1948).

BIOLOGICAL APPLICATIONS

70. Radin, N. S., *Nucleonics*, **1,** No. 1, 24; No. 2, 48; No. 4, 51 (1947);
     2, No. 1, 50; No. 2, 33 (1948).
71. Sacks, J., *Chem. Revs.*, **42,** 411 (1948).
72. Calvin, M., *Nucleonics*, **2,** No. 3, 40 (1948).
73. Rabinowitch, E. I., *Photosynthesis and Related Processes*, Vol. I. Inter-
     science, New York, 1944. Vol. II, in preparation.
74. Wood, H. G., *Physiol. Revs.*, **26,** 198 (1946).
75. Buchanan, J. M., and A. B. Hastings, *Physiol. Revs.*, **26,** 120 (1946).
76. Gurin, S., and A. M. Delluva, *Federation Proc.*, **6,** 257 (1947).
77. Quimby, E. H., *Nucleonics*, **1,** No. 4. 2 (1947).
78. Greenberg, D. M., R. B. Aird, M. D. D. Boelter, W. W. Campbell,
     W. E. Cohn, and M. M. Murayama, *Am. J. Physiol.*, **140,** 47 (1943).
79. Melchior, J. B., and H. Tarver, *Arch. Biochem.*, **12,** 301, 309 (1947).
80. Hahn, P. F., L. L. Miller, F. S. Robscheit-Robbins, W. F. Bale, and G.
     H. Whipple, *J. Exptl. Med.*, **80,** 77 (1944).
81. Moore, G. F., *Science*, **107,** 569 (1948).

COMMERCIAL SOURCES

82. National Technical Laboratories, So. Pasadena, Calif.
83. Fred C. Henson Co., Pasadena, Calif.
84. Instrument Development Laboratories, Chicago, Ill.
85. Radiation Counter Laboratories, Chicago, Ill.
86. Able Scientific Company, Chicago, Ill.
87. Victoreen Instrument Company, Cleveland, Ohio.
88. Tracer-lab, Inc., Boston, Mass.
89. Electronics, Inc., Philadelphia, Penna.
90. Potter Instrument Co., Flushing, N. Y.
91. Berkeley Scientific Co., Richmond, Calif.
92. Kelley-Koett Manufacturing Co., Covington, Ky.
93. Rauland Corporation, Chicago, Ill.
94. Sylvania Electric Products, Inc., New York, N. Y.
95. Texas Research Foundation, Renner, Texas.
96. Sorenson and Co., Stamford, Conn.
97. Abbott Laboratories, North Chicago, Ill.

SUBJECT INDEX

A

Aberrations, lenses, 347, 349–50, 374

Abramson cell, 282–284

Absorption, self-absorption by radio-isotopes, 627–30
  X-ray, 458–61, 465–6

Absorption coefficient, 423–5. See also *Extinction coefficient.*

Absorption estimation (light), 438–45

Absorption laws (light), 422–5

Absorption spectra, carotinoid pigments, 420
  7-dehydrocholesterol, 447
  light, 420–30
  living systems, 427–8
  measurements, 425–7
  nucleic acid, 446
  porphyrins, 420
  protein, 446
  rhodopsin, 433
  scattering materials, 428–30

Absorption spectrophotometry, 426–7

Abundance ratio (isotopes), 562, 589

Accuracy, isotope measurements, 586–9, 623–30
  research, 29

Acetone, mass spectrum, 570

Achromatic objectives, 349–51

Acoustic resistance, 303

Acoustics, 301–42. See also *Ultrasonic vibrations.*

Action potentials, 259–62

Action spectra, 431–47
  antirachitic action, rats, 447
  interpretation, 440–7
  killing bacteria, 431, 446
  killing fungus, 446
  various biological phenomena, 431

Adiabatic calorimeters, 184

Adiabatic method (calorimetry), 186–7

Afterperiod (calorimetry), 194–5

Airy disc, 345

Alpha particles, 491 ff. See also *Ionizing radiations.*
  dose rate, 535, 545–6
  ion density, 502–4
  sources, 535–7

Alpha rays. See *Alpha particles.*

Aluminum, penetrability by beta rays, 602–3

Amoeba, viscosity studies, 120, 126

Amplifier(s), capacitor-coupled, 254–5
  direct current, 250–4
  electrometer, 249–50
  electronic, 248–58
  limitations, 255–8
  in mass spectrometry, 575–7
  power supplies, 257–8

Angstrom unit, 383

Animal calorimetry, 200–3

Aperture, numerical, 345–7, 410

Apochromatic objectives, 350–1

Arcs, carbon, 373
  mercury, 352, 433–4
  zirconium, 373

Artifacts, desiccation, 410–12

Attenuation coefficient, 430

Autoradiograph, 369, 636–8

B

Background count (radioactivity), 624–5

Bacteria, preservation, 215, 228–9

Bacteriophage, preservation, 216

Beckmann thermometer, 136, 139

Beer's law, 423–4
  deviations from, 428

Bernstein's hypothesis, 264

Beryllium targets, 528–31

Beta particles. See also *Electrons* and *Beta rays.*
  mean energy, 524–5

Beta rays, absorption in aluminum, 602–3
  danger from, 497, 521
  penetrating power, 520–2

Energy, distribution, beta rays, 524–5
  equivalents (table), 180
  excitation, 557
  ionization, 551
  unit, 540
Environmental control, 26–27
Enzyme, inactivation, 419, 514, 516
Epi-illuminator, 357–8
Equivalent circuits, 235–7, 264–5
  living cells, 264–5
Errors, classification, 31–32
  instrumental, 203–4
  interpretation, 205
  probable, 624
  radioisotope measurements, 623–30
  sampling, 204, 626–7
  self-absorption, 627–30
Erythema, dose, 456
  sunburn, 441–2
Escherichia coli, 499–500, 505–9
Exchange reactions, radioisotopes, 632
  stable isotopes, 568, 591–2
Excitation, energy, 557
Experimental data, accuracy, 29
  analysis, 29–32
  choice of units, 27–28
  environmental control, 26–27
  irradiation studies, 470–3
Experimental limitations, instrumentation and technique, 17–18
  theoretical considerations, 15–17
Experiments, design, 25–26, 37
  fruitless, 1–38
  negative, 15
  timing, 23–25
Extinction coefficient, 423–4
  molecular, 424
Eyepiece, 355–6, 374–5. See also *Oculars.*

F

Facts, determined by vote, 32
Film badge technique, 641–2
Filters, color, 350–3, 364, 434
  color-compensating, 350
  interference, 434
  Polaroid, 362, 374
  X-ray, 461–2

Fixation, by freezing-drying, 218, 223
  nitrogen, 595
Flow method, in calorimetry, 189–93
Fluorescence, 349, 363–4, 430
Fluorescent specimens, 363–5
Fluorite objectives, 350
Fourier's law, 187
Freezing, biological materials. See *Freezing-drying process* and *Quick-freezing.*
Freezing-drying process, accomplishments, 226–31
  advantages, 212–4
  applications, 214–9
  equipment, 218–21
  limitations, 221–6
  technique, 221–6
Freezing point depression, 56–7
Frictional coefficient, 84, 88
Frictional ratio, significance, 94
Fürth formula, 124

G

Galvanometers, 256–8
Galvanometric method (thermocouples), 144–5
Gamma rays, 512. See also *Electrons* and *Ionizing radiations.*
  absorption in hydrogen, 555
  absorption in lead, 602–3
Gas analysis, mass spectrometry, 570–1
Gaussian distribution law, 623–4
Geiger counters, background counts, 624–6
  end window type, 619–23
  gas refilling, 618–20
  principle, 616–8
  quenching circuits, 617–8
  plateau curve, 617–8
Geiger-Mueller counter. See *Geiger counters.*
"GM tube," 643. See also *Geiger counters.*
Gibbs-Helmholtz equation, 47
Glass, heat-absorbing, 373
Glycerol, immersion fluid, 364
Grid current–voltage curve, 251
Grotthus-Draper law, 418